

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平6-282385

(43)公開日 平成6年(1994)10月7日

(51)Int.Cl.⁵

G 0 6 F 3/06

識別記号

庁内整理番号

F I

技術表示箇所

3 0 5 F 7165-5B

3 0 1 Z 7165-5B

審査請求 未請求 請求項の数 9 O L (全 32 頁)

(21)出願番号 特願平5-67017

(22)出願日 平成5年(1993)3月25日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 加賀 隆志

神奈川県川崎市麻生区王禅寺1099 株式会

社日立製作所システム開発研究所内

(72)発明者 佐藤 孝夫

神奈川県川崎市麻生区王禅寺1099 株式会

社日立製作所システム開発研究所内

(72)発明者 山本 彰

神奈川県川崎市麻生区王禅寺1099 株式会

社日立製作所システム開発研究所内

(74)代理人 弁理士 富田 和子

最終頁に続く

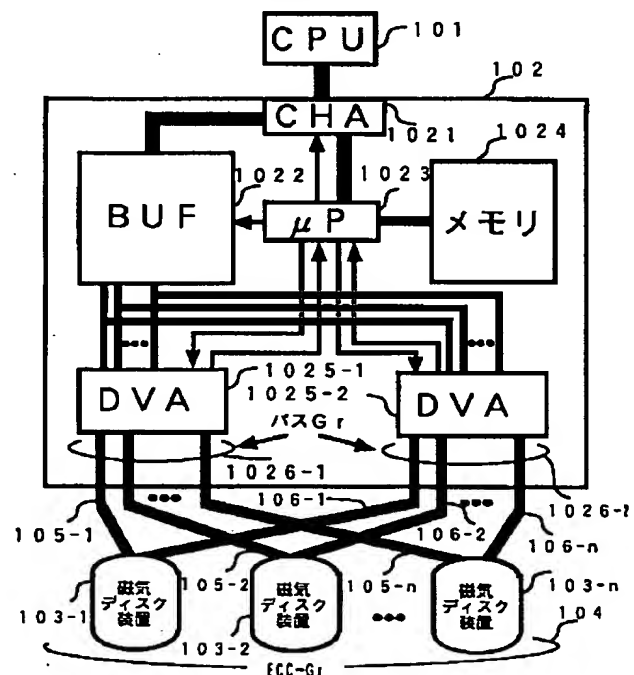
(54)【発明の名称】 記憶制御装置および記憶制御装置を備える情報処理システム

(57)【要約】 (修正有)

【目的】記憶制御装置内のバスを多重に備え、複数の記憶装置を並列制御し、信頼性の向上を図る。

【構成】情報を分割して誤り訂正用の情報を生成し、分割した情報とともに転送路を介して複数の記憶手段103-1~103-nに書き込む制御、並びに読み出し時に前記分割した情報及び誤り訂正用の情報に基づいて情報の誤り検出を行い、誤りがあった場合には情報を回復する記憶制御装置であって、複数の記憶手段との間で第1の転送路群を介して情報の転送を行う第1の転送手段1025-1、及び第2の転送路群を介して情報の転送を行う第2の転送手段1025-2とを備え、双方の転送路は、転送路が故障していることを検出する検出手段、故障している転送路を使用しないように閉塞状態にする閉塞手段、及び故障している転送路に接続されている記憶手段に接続される他の転送路群の転送路を使用しないように論理的に閉塞状態にする論理閉塞手段を備える。

図1 本発明の一実施例



1

【特許請求の範囲】

【請求項1】情報を分割し、該情報に基づいて誤り訂正用の情報を生成し、前記分割した情報と該誤り訂正用の情報とを転送路を介して複数の記憶手段に書き込む制御と、該複数の記憶手段からの読み出しの制御とを行い、読み出し時に前記分割した情報と該誤り訂正用の情報とに基づいて情報の誤り検出を行い、誤りがあった場合には情報を回復する記憶制御装置であって、

前記複数の記憶手段との間で第1の転送路群を介して情報の転送を行う第1の転送手段と、前記複数の記憶手段との間で第2の転送路群を介して情報の転送を行う第2の転送手段とを備え、

該第1の転送手段および第2の転送手段の各々は、前記転送路が故障していることを検出する検出手段と、該検出手段で検出した故障している転送路を使用しないように閉塞状態にする閉塞手段と、

該検出手段で検出した故障している前記転送路に接続されている記憶手段に接続されている、他の転送路群の転送路を使用しないように論理的に閉塞状態にする論理閉塞手段とを備えることを特徴とする記憶制御装置。

【請求項2】情報を分割し、該情報に基づいて誤り訂正用の情報を生成し、前記分割した情報と該誤り訂正用の情報とを転送路を介して複数の記憶手段に書き込む制御と、該複数の記憶手段からの読み出しの制御とを行い、読み出し時に前記分割した情報と該誤り訂正用の情報とに基づいて情報の誤り検出を行い、誤りがあった場合には情報を回復する記憶制御装置であって、

前記複数の記憶手段との間で第1の転送路群を介して情報の転送を行う第1の転送手段と、前記複数の記憶手段との間で第2の転送路群を介して情報の転送を行う第2の転送手段とを備え、

該第1の転送手段および第2の転送手段の各々は、前記転送路が故障していることを検出する検出手段と、該検出手段で検出した故障している転送路を使用しないように閉塞状態にする閉塞手段とを備え、

前記情報の転送時に、前記検出手段で検出した転送路に接続されている記憶手段に接続されている転送手段を読み出し専用とすることを特徴とする記憶制御装置。

【請求項3】請求項2において、前記情報の転送時に、前記検出手段で検出した転送路に接続されている記憶手段に接続されている他の転送路群の転送路を使用しないように論理的に閉塞状態にする論理閉塞手段をさらに備えることを特徴とする記憶制御装置。

【請求項4】請求項1または3において、前記論理的閉塞状態を使用可能な状態にする閉塞解除手段をさらに備えることを特徴とする記憶制御装置。

【請求項5】情報を分割し、該情報に基づいて誤り訂正用の情報を生成し、前記分割した情報と該誤り訂正用の情報とを転送路を介して複数の記憶手段に書き込む制御と、該複数の記憶手段からの読み出しの制御とを行い、

2

読み出し時に前記分割した情報と該誤り訂正用の情報とに基づいて情報の誤り検出を行い、誤りがあった場合には情報を回復する記憶制御装置であって、

前記複数の記憶手段との間で第1の転送路群を介して情報の転送を行う第1の転送手段と、前記複数の記憶手段との間で第2の転送路群を介して情報の転送を行う第2の転送手段とを備え、

該第1の転送手段および第2の転送手段の各々は、

前記転送路が故障していることを検出する検出手段と、該検出手段で検出した故障している転送路を使用しないように閉塞状態にする閉塞手段と、

全ての転送路のうち、前記記憶手段に対応する使用可能な転送路を選択して転送路群とし、該選択された転送路群を介して情報を転送する選択転送手段とを備えることを特徴とする記憶制御装置。

【請求項6】請求項5において、前記第1の転送手段と第2の転送手段と前記選択転送手段とのうち、閉塞状態にある前記転送路の数がより少ないものを利用して転送することを特徴とする記憶制御装置。

【請求項7】請求項5または6において、情報の転送時に、前記検出手段で検出した転送路に接続されている記憶手段に接続されている前記第1の転送手段または第2の転送手段を読み出し専用とすることを特徴とする記憶制御装置。

【請求項8】請求項1、2または5において、前記転送路が閉塞状態か否かの状態を示す情報を格納する格納手段を備え、転送時に該格納手段を参照して転送の制御をすることを特徴とする記憶制御装置。

【請求項9】情報の処理をするホスト計算機と、情報を記憶する複数の記憶装置と、該記憶装置の読み書きの制御を行う記憶制御装置とを備える情報処理システムにおいて、

前記複数の記憶装置の各々と前記記憶制御装置とは、第1および第2の転送路群を介して接続され、

前記記憶制御装置は、前記ホスト計算機における情報を分割し、該情報に基づいて誤り訂正用の情報を生成し、前記分割した情報と該誤り訂正用の情報とを前記転送路を介して前記記憶装置に対して書き込む制御を行い、該複数の記憶手段からの読み出しの制御を行い、読み出し時に前記分割した情報と該誤り訂正用の情報とに基づいて情報の誤り検出を行い、誤りがあった場合には情報を回復する制御部と、

前記複数の記憶手段との間で前記第1の転送路群を介して情報の転送を行う第1の転送手段と、

前記複数の記憶手段との間で前記第2の転送路群を介して情報の転送を行う第2の転送手段とを有し、

該第1の転送手段および第2の転送手段の各々は、

前記転送路が故障していることを検出する検出手段と、該検出手段で検出した故障している転送路を使用しないように閉塞状態にする閉塞手段と、

10

20

30

40

50

該検出手段で検出した故障している転送路に接続されている記憶装置に接続されている、他の転送路群の転送路を使用しないように論理的に閉塞状態にする論理閉塞手段とを備え、

前記記憶制御装置は、各転送路が閉塞状態か否かの状態を示す情報を格納する格納手段を備え、

前記制御部は、転送時に、該格納手段を参照し、第1の転送手段および第2の転送手段に対して転送の制御をすることを特徴とする情報処理システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、ディスクアレイを構成する記憶装置を、複数のバスで制御する記憶制御装置に関する。

【0002】

【従来の技術】記憶制御装置の性能と信頼性を向上する技術として、エー・シー・エム シグモッド 88 (1988年)の第109頁から第116頁の論文「ア ケ イ ス フ ォー リ ダ ン ダ ン ト ア レ イ ズ オ ブ イ ン エ ク ス ペ ン シ ブ ディ ス ク ス (ACase for Redundant Arrays of Inexpensiv Disks(RAID))」(David A. Patterson, Garth Gibson and Randy H. Katz)においてRAID(Redundant Arrays of Inexpensive Disks)と呼ばれる技術が紹介されている。

【0003】上記RAIDの特徴は、複数の記憶装置を並列制御することにより応答性能またはスループットを従来の記憶制御装置より高めることにある。また、記憶装置毎にデータを選び、それらデータに対する誤り訂正符号(以下ECC)に代表される冗長データを格納する記憶装置を設ける。このため、故障した記憶装置の個数が冗長データで訂正可能な個数ならば、故障した記憶装置のデータを失わない。つまり、RAIDの他の特徴は信頼性が優れていることにある。

【0004】前記論文では、RAIDは5つのグループに分類されている。その中の2番目(以下、RAID2という)と、3番目(以下、RAID3という)とは、データレコードを分割して複数の記憶装置と一斉にデータ転送をおこなうので応答性能に優れている。RAID2とRAID3との違いは、RAID2がECCにハミング符号を使うのに対し、RAID3はパリティを用いることにある。

【0005】一方、記憶制御装置の信頼性を向上することを目的に、記憶制御装置内の構成要素を多重に備えることが従来とられている。例えば、アイビーエム システムズ ジャーナル、Vol. 28、No. 2 (1989年)第196頁から第226頁(IBM SYSTEMS JOURNAL, VOL. 28, No. 2, 1989)において、1個のディスクボリュームに対し2個のバスでデータ転送できる3880モデル3と、1個のディスクボリュームに対し4個のバスでデータ転送できる3990モデル2とが紹介されてい

る。

【0006】

【発明が解決しようとする課題】制御装置内のバスを多重に持つ上記技術を、上記RAID2またはRAID3に適用する場合について考える。前記2つの技術は、共に記憶制御装置の信頼性を向上させるので、個々の技術のみを用いるよりさらに信頼性の向上が期待できる。

【0007】ここで、データレコードを分割してECCを付加した結果の複数のデータを格納する記憶装置の集合体をECCグループ(以下、ECC-Gr)と呼ぶ。また、ECC-Grの個々の記憶装置に接続されているデータ転送路(以下、バス)の集合体をデータ転送路グループ(以下、バスGr)と呼ぶ。バスGrが複数ある時、個々の記憶装置に対するバスの故障が独立に発生しうするために、バスGr毎に転送不可能な記憶装置が異なる場合がある。転送不可能な記憶装置が存在するバスGrで書き込みをすると、古いデータが残る記憶装置ができる。この時、この記憶装置と転送可能なバスGrでECC-Grからの読み出しをおこなうと、この記憶装置の古いデータを読み出してしまふ。記憶制御装置は、そのデータが古いということを認識できないので、古いデータを含んだデータレコードを正しいものとしてCPUに転送してしまふ。また、読み出したデータを必ずECCでチェックするとしても、記憶装置としては故障していないので、ECCがパリティであるRAID3では誤り位置が分からず訂正ができない。つまり、書き込みに使用したバスGrと転送不可能な記憶装置が一致するバスGrで読み出さないとデータの訂正ができない。

【0008】以上のように、あるバスGrが使用不可能になると、そのバスGrを使用して書き込んだデータレコードを正しく読み出せなくなる。つまり、従来のように記憶制御装置の信頼性の向上を目的としてバスGrを複数個設けても従来通りの効果を望めない。

【0009】例えば、図2を参照して課題を詳細に説明する。図2において、複数の磁気ディスク装置103-1~103-nと記憶制御装置1025-1および1025-2とを備えている。記憶制御装置と磁気ディスク装置間は、書き込みに使用するバス105-1~105-nのバスGr1026-1と、読み込みに使用するバス106-1~106-nのバスGr1026-2とがある。図2に示すような構成において、バス105-1が故障している状態でバスGr1026-1でデータレコードを書き込んだ場合、磁気ディスク装置103-1には書き込みが行われず、古いデータが残る。この状態でバスGr1026-2でデータレコードを読み出すと、磁気ディスク装置103-1から古いデータを読み出してしまふ。したがって、不正なデータレコードをCPU101に転送してしまふ。また、バスGr1026-1が使用不可能になると、バスGr1026-2を構成するバスが全く故障していなくても、バスGr102

6-1で書き込んだデータレコードを正しく読み出せなくなる。

【0010】さらに、前記論文「ア ケイ ス フォー リ ダ ン ダ ント ア レ イ ズ オ ブ イ ン エ ク ス ペ ン シ ブ デ ィ ス ク ス」には複数のパスGrを使用してRAID2およびRAID3を制御する場合のことは言及されていないので、どのように制御を行うかは考慮されていない。

【0011】そこで、本発明の目的は、記憶制御装置内のパスを多重に備え、複数の記憶装置を並列制御し、信頼性の向上を図ることができる記憶制御装置を提供する。

【0012】

【課題を解決するための手段】本発明は、上記課題を解決するために、情報を分割し、該情報に基づいて誤り訂正用の情報を生成し、前記分割した情報と該誤り訂正用の情報とを転送路を介して複数の記憶手段に書き込む制御と、該複数の記憶手段からの読み出しの制御とを行い、読み出し時に前記分割した情報と該誤り訂正用の情報とに基づいて情報の誤り検出を行い、誤りがあった場合には情報を回復する記憶制御装置であって、前記複数の記憶手段との間で第1の転送路群を介して情報の転送を行う第1の転送手段と、前記複数の記憶手段との間で第2の転送路群を介して情報の転送を行う第2の転送手段とを備え、該第1の転送手段および第2の転送手段の各々は、前記転送路が故障していることを検出する検出手段と、該検出手段で検出した故障している転送路を使用しないように閉塞状態にする閉塞手段と、該検出手段で検出した故障している前記転送路に接続されている記憶手段に接続されている、他の転送路群の転送路を使用しないように論理的に閉塞状態にする論理閉塞手段とを備える。この場合、前記論理的閉塞状態を使用可能な状態にする閉塞解除手段をさらに備えるようにしてもよい。

【0013】また、記憶制御装置は、前記情報の転送時に、前記検出手段で検出した転送路に接続されている記憶手段に接続されている転送手段を読み出し専用とすることができる。

【0014】さらに、全ての転送路のうち、前記記憶手段に対応する使用可能な転送路を選択して転送路群とし、該選択された転送路群を介して情報を転送する選択転送手段とを備えることができる。また、前記第1の転送手段と第2の転送手段と前記選択転送手段とのうち、閉塞状態にある前記転送路の数がより少ないものを利用して転送することができる。この場合において、記憶制御装置は、情報の転送時に、前記検出手段で検出した転送路に接続されている記憶手段に接続されている前記第1の転送手段または第2の転送手段を読み出し専用とすることができる。また、前記転送路が閉塞状態か否かの状態を示す情報を格納する格納手段を備え、転送時に該

格納手段を参照して転送の制御をすることができる。

【0015】また、上記記憶制御装置を、情報の処理をするホスト計算機と、情報を記憶する複数の記憶装置とを備える情報処理システムに適用してもよい。

【0016】

【作用】本発明の記憶制御装置は、情報を分割し、該情報に基づいて誤り訂正用の情報を生成し、前記分割した情報と該誤り訂正用の情報(ECC)とを転送路(パス)を介して記憶手段に書き込む制御と、該複数の記憶手段からの読み出しの制御とを行い、読み出し時に前記分割した情報と該誤り訂正用の情報とに基づいてデータの誤り検出を行い、誤りがあった場合にはデータを回復する。

【0017】また、記憶制御装置は、前記複数の記憶手段との間で第1の転送路群(パスGr1)を介して情報の転送を行う第1の転送手段と、前記複数の記憶手段との間で第2の転送路群を介して情報の転送を行う第2の転送手段(パスGr2)とを備える。該第1の転送手段および第2の転送手段の各々の検出手段は、前記転送路が故障していることを検出する。閉塞手段は、該検出手段で検出した故障している転送路を使用しないように閉塞状態にする。論理閉塞手段は、該検出手段で検出した故障している前記転送路に接続されている記憶手段に接続されている、他の転送路群の転送路を使用しないように論理的に閉塞状態にする。すなわち、故障したパスに接続されている記憶手段に対しては、読み書き込みをしないように、該記憶手段に接続されているパスを擬似的に故障とする。このようにすると、古いデータを読んでしまう事を防ぐことができる。

【0018】また、記憶制御装置は、情報の転送時に、前記検出手段で検出した転送路に接続されている記憶手段に接続されている第1または第2の転送手段を読み出し専用とするようにしてもよい。これにより、故障をしたパスを含んでいる第1または第2の転送手段において、記憶手段に情報を書き込むことがない。また、書き込み時に使用したパスGrが使用不可能になっても、他のパスGrを使用してデータレコードとしての読み出しができる。

【0019】また、前記閉塞状態を使用可能な状態にする閉塞解除手段をさらに備える場合には、例えば、前記閉塞手段により閉塞状態にした後で、故障したパスを含むパスGrが読み出し専用にしたときに、閉塞解除手段により閉塞を解除することができる。

【0020】また、選択転送手段を備える場合には、選択転送手段は、全ての転送路のうち、前記複数の記憶手段に対応する使用可能な転送路を選択して情報を転送する。前記第1の転送手段および第2の転送手段においてパスの故障が多い場合などには、選択転送手段により全てのパスの中から使用可能なパスのみを選択して転送できる。

【0021】また、前記第1の転送手段と第2の転送手段と前記選択転送手段とのうち（これらを組合せパスGrという）、閉塞状態である転送路の数がより少ないものを利用して転送する。この場合には、組合せパスGrのうち、閉塞状態がより少ないものを利用して読み書きを行うことができる。また、読み出し専用や単純閉塞の機能を併せて備えてもよい。

【0022】以上より、記憶制御装置の信頼性が向上する。

【0023】

【実施例】図1は、本発明の一実施例に基づく記憶制御装置を示す図である。図1において、本発明の記憶制御装置102は、チャンネルアダプタ（以下CHA）1021と、データを一時的に保存できるバッファ（以下BUF）1022と、 n ($2 \leq n$) 個の磁気ディスク装置と一斉にデータ転送をおこなえる、第1および第2の転送手段のデバイスアダプタ（以下DVA）1025-1および1025-2と、 n 個の磁気ディスク装置103-1、103-2、 \dots 、103- n と、DVA1025-1と n 個の磁気ディスク装置とを接続する n 個のパス105-1、105-2、 \dots 、105- n と、DVA1025-2と n 個の磁気ディスク装置とを接続する n 個のパス106-1、106-2、 \dots 、106- n と、制御部のマイクロプロセッサ（図中 μP で表記）1023と、メモリ1024とを備える。

【0024】ここで、CHA1021とBUF1022とは1個のパスで接続している。また、BUF1022と各DVA1025-1および1025-2とは、 n 個のパスで接続している。また、マイクロプロセッサ1023は、メモリ1024に格納されているプログラムや各種制御用のデータに従い、本実施例における記憶制御装置の制御をおこなう。CHA1021は、CPU101とマイクロプロセッサ1023間のデータ転送をする。そして、マイクロプロセッサ1023の指示によって、CHA1021はCPU101とBUF1022間のデータの転送をする。また、DVA1025-1および1025-2はマイクロプロセッサ1023の指示によってBUF1022と n 個の磁気ディスク装置103-1、103-2、 \dots 、103- n 間のデータ転送を一斉におこなう。また、DVA1025-1および1025-2は、BUF1022と個々の磁気ディスク装置間のデータ転送の失敗を検出し、マイクロプロセッサ1023にデータ転送の失敗があったことを報告する。また、DVA1025-1および1025-2は、4台以上の磁気ディスク装置が読み出し可能ならば、1台が故障していても故障したディスク装置に格納しているデータを回復することができる。また、磁気ディスク装置は、データを格納するディスク装置と、誤り訂正符号を記憶するディスク装置とを備える。また、DVA1025-1と n 個の磁気ディスク装置とを接続する n 個のパス

105-1、105-2、 \dots 、105- n をパスGr1とする。DVA1025-2と n 個の磁気ディスク装置とを接続する n 個のパス106-1、106-2、 \dots 、106- n をパスGr2とする。すなわち、ディスク装置に接続するパスを、パスGr1とパスGr2とにより2重化している。また、さらに、パスGrおよびDVAを設けるようにしてもよい。ここで、データレコードを分割して誤り訂正符号（ECC）を付加した結果の複数のデータを格納する記憶装置の集合体をECCグループ（以下、ECC-Gr）と呼ぶ。ECCグループの磁気ディスク装置のうち、2個が誤り訂正符号（ECC）を記憶するものであり、他の $(n-2)$ 個の磁気ディスク装置はデータを記憶する。そして、このECCは、 n 個の磁気ディスク装置の内、磁気ディスク装置の故障によって、任意の2個までの磁気ディスク装置から読み出しができないときに、読み出せないデータを生成するための符号である。

【0025】また、上記図1に示す記憶制御装置を備える情報処理システムのハードウェア構成を図32に示す。図32において、情報処理システムは、情報の処理をするホスト計算機と、情報を記憶する複数の記憶装置と、該記憶装置の読み書きの制御を行う記憶制御装置とを備える。記憶制御装置は、前記ホスト計算機における情報を分割し、該情報に基づいて誤り訂正用の情報を生成し、前記分割した情報と該誤り訂正用の情報とを前記転送路を介して前記記憶装置に対して書き込む制御を行い、該複数の記憶手段からの読み出しの制御を行い、読み出し時に前記分割した情報と該誤り訂正用の情報とに基づいて情報の誤り検出を行い、誤りがあった場合には情報を回復する制御部と、前記複数の記憶手段との間で前記第1の転送路群を介して情報の転送を行う第1の転送手段と、前記複数の記憶手段との間で前記第2の転送路群を介して情報の転送を行う第2の転送手段とを有し、該第1の転送手段および第2の転送手段の各々は、前記転送路が故障していることを検出する検出手段と、該検出手段で検出した故障している転送路を使用しないように閉塞状態にする閉塞手段と、該検出手段で検出した故障している転送路に接続されている記憶装置に接続されている、他の転送路群の転送路を使用しないように論理的に閉塞状態にする論理閉塞手段とを備え、前記記憶制御装置は、各転送路が閉塞状態か否かの状態を示す情報を格納する格納手段を備え、転送時に、該格納手段を参照し、第1の転送手段および第2の転送手段に対して転送の制御をする。

【0026】上記記憶制御装置において、第1の実施例について図26、図27および図31を参照して説明する。第1の実施例を説明するための模式図を図26、図27および図31に示す。第1の実施例では、図31に示すように、ディスク装置に対して読み書きを行う際に、パスが使用不可能な状態を検出手段により検出し、

不可能な場合には、そのパスに接続しているディスク装置に読み書きを行わないようにするために該ディスク装置に接続している他のパスに対して論理閉塞手段により論理閉塞を行うように制御する。このように、論理閉塞を行うことにより全てのパスGr間で入出力可能なディスク装置を一致させる。また、本実施例においては、図26および図27に示すように、使用不可能なパスがある場合に、該使用不可能なパスを含むパスGrを読み出し専用とする場合について併せて説明している。

【0027】以下に、メモリ1024に格納している内容を説明した後に、全てのパスGr間で入出力可能なディスク装置を一致させる場合と、一方のパスGrを読み出し専用にする場合の記憶制御装置の処理動作を説明する。

【0028】図3を用いてメモリ1024に格納される内容を説明する。メモリ1024は、書き込み処理用のプログラムの領域301と、読み出し処理用のプログラムの領域302と、パスGr機能状態決定プログラムの領域303と、パスの状態の領域304と、パスGrの機能状態の領域305とに分かれている。書き込み処理用のプログラムの領域301には、磁気ディスク装置にデータを書き込む際の処理用のプログラムが格納されている。読み出し処理用のプログラムの領域302には、

磁気ディスク装置からデータを読み出す際の処理用のプログラムが格納されている。書き込みおよび読み出しの処理フローについては後述する。パスGr機能状態決定プログラムの領域303には、パスGrが書き込みおよび読み出しが可能か、否か、または、読み出しのみ可能かの状態を決定するためのプログラムが格納されている。パスの状態の領域304には、各パスごとの使用可能か否かと、論理閉塞されているかを示す状態を格納する。パスGrの機能状態の領域305は、パスGr機能状態決定プログラムにより決定されてパスGrの機能状態を格納する。

【0029】パスの状態の領域304における各パスの状態について、図4および表1を参照して説明する。領域304は、図1におけるパス105-1、105-2、・・・、105-nの状態をそれぞれ表す、領域401-1、401-2、・・・、401-nと、図1におけるパス106-1、106-2、・・・、106-nの状態をそれぞれ表す、領域402-1、402-2、・・・、402-nとに分かれている。パスの状態には、表1に示す正常、論理閉塞、故障の3状態がある。

【0030】

【表1】

表1 パスの状態

項番	状態名	内容	意味
1	正常	11	使用可能
2	論理閉塞	10	物理的には正常だが、使用不可能
3	故障	00	物理的に故障していて、使用不可能

【0031】表1において、正常は、そのパスがBUF1022とそのパスが接続する磁気ディスク装置間のデータ転送に使用可能であることを示す。論理閉塞は、そのパスが、物理的には正常であってBUF1022とそのパスが接続する磁気ディスク装置間のデータ転送に使用可能だが、普段の入出力には使用不可能として扱う。すなわち、論理的に閉塞状態にするときにこの状態に設定する。ただし、磁気ディスク装置内のデータを回復するときには、書き込みのみ使用可能である。故障は、そのパスが、物理的な故障が原因でBUFと磁気ディスク装置間のデータ転送に使用不可能であることを示す。

正常を示す値は2進数の11、論理閉塞を示す値は2進数の10、故障を示す値は2進数の00である。2桁の2進数の上位が1なら物理的に正常であることを示し、0なら物理的に故障中であることを示している。2桁の2進数の下位が1なら、制御上正常に扱うことを示し、0なら制御上故障として扱うことを示している。

【0032】また、パスGrの機能状態の領域305を図5および表2を参照して説明する。

【0033】

【表2】

11
表2 ハスGrの機能状態
12

項番	状態名	内容	意 味
1	正常	11	書き込みと読み出しに使用可能
2	読み出し専用	10	読み出しのみに使用可能
3	閉塞	00	書き込みと読み出しに使用不可能

【0034】領域305は、パスGr1026-1と、1026-2との機能状態を表すための領域として、領域501と502とに分かれている。パスGrの機能状態には表2に示す正常、読み出し専用および閉塞の3状態がある。表2において、正常は、そのパスGrが書き込みと読み出しに使用可能であることを示す。読み出し専用は、そのパスGrが読み出しのみに使用可能であることを示す。閉塞は、そのパスGrが書き込みと読み出しに使用不可能であることを示す。正常を示す値は2進数の11、読み出し専用を示す値は2進数の10、閉塞を示す値は2進数の00である。2桁の2進数の上位が1なら読み出しに使用可能であることを示し、0なら読み出しに使用不可能であることを示している。2桁の2進数の下位が1なら書き込みに使用可能であることを示し、0なら書き込みに使用不可能であることを示している。

【0035】次に、図8を用いてパスGr機能状態決定のフローを説明する。図8に示すフローは、前述の図3に示すメモリ1024に格納されるパスGr機能状態決定プログラムのフローを示している。

【0036】図8において、マイクロプロセッサ1023は、メモリ1023の領域304を参照して故障または論理閉塞のパスを数える。メモリ1023の領域304に格納されているパス状態は、DVA1025-1および1025-2によりアクセス時に故障であることが報告され、マイクロプロセッサ1023によりそれ以前に設定されている。故障または論理閉塞のパスが3個以上あるパスGrは、パスGr全体を閉塞にする(処理801)。ここで、3個以上としたのは、RAIDを利用したパスでは、ECCにより任意の2個までの磁気ディスク装置から読み出しができないときに、読み出せないデータを生成することができるが、3個以上の場合にはデータを生成できないからである。3個以上故障している場合には、そのデータを復元することができないので、パスGr全体を閉塞して利用しないようにする。処理801において閉塞されるパスGrが生じたら(処理802)、閉塞されていないパスGrを正常として決定し、パスGr機能状態決定処理を終了する(処理803)。閉塞されているパスGrがなければ、パスGrに閉塞がない時のパスGr機能状態決定処理をおこなう

(処理804)。

【0037】以下に、パスGrに閉塞がない時のパスGr機能状態決定処理のフローを図9を参照して説明する。図9において、マイクロプロセッサ1023は、メモリ1024の領域401-1と領域402-1とに格納されているパス状態の内容が一致するか否かを調べる。同様に、領域401-2と領域402-2とに格納されているパス状態の内容、・・・、領域401-nと領域402-nとに格納されているパス状態の内容それぞれが一致するかを調べる。つまり、入出力可能な磁気ディスク装置がパスGr間で完全に一致するか否かを調べる(処理901)。完全に一致するのなら2個のパスGrを共に正常にし、パスGr機能状態決定処理を終了する(処理902)。完全には一致しないならば、故障のパスの個数に大小関係があるかを調べる(処理903)。大小関係があれば、故障の多いパスGrを読み出し専用にし、少ないパスGrを正常にする(処理904)。大小関係がなければ、現在のパスGr機能状態を保持する(処理905)。次に、正常のパスGrにおいて、故障しているパスに接続している磁気ディスク装置に対して読み書きを行わないようにするために、図27に示すように、該ディスク装置に接続している他の正常なパスを論理閉塞(すなわち、擬似故障)にパス状態を設定する(処理906)。そして最後に、故障または論理閉塞のパスが3個以上のパスGrを閉塞にする(処理907)。

【0038】以上のように、パスGrの機能状態の決定処理は、記憶制御装置を初期作動させる時と、DVAがパスの故障をマイクロプロセッサ1023に報告した時にマイクロプロセッサ1023がおこなう。

【0039】次に、記憶制御装置における磁気ディスク装置へのデータの書き込み処理と読み出し処理との処理のフローを説明する。

【0040】まず、図6を参照して書き込み処理のフローを説明する。図6に示す書き込み処理のフローは、前述の図3に示すメモリ1024に格納される書き込み処理用のプログラムのフローを示している。

【0041】図6において、マイクロプロセッサ1023は、書き込むデータレコードをCHA1021からBUF1022へ転送する(処理601)。次に、使用す

るパスGrを選択する。この処理は、メモリ1024の領域305に記憶している全てのパスGrの機能状態を参照して正常のパスGrを探す。正常のパスGrが複数個あれば、使用頻度の低いものを選択する。正常のパスGrがなければ、CPUに異常終了を報告し、書き込み処理を終了する(処理602、603)。使用するパスGrを選択できたら、BUF1022に転送された書き込むデータレコードを各磁気ディスク装置に分割格納するために(n-2)等分する(処理604)。次に、この(n-2)個のデータの集まりに対してECCを作成し、データをn個にする(処理605)。これらn個のデータは、それぞれの磁気ディスク装置に書き込まれる。ECCを作成した後に、選んだパスGrが接続するDVAを選択する(処理606)。n個のデータをBUF、DVA、磁気ディスク装置の順に転送する(処理607)。転送に成功したらCPUに正常終了を報告し、失敗したらCPUに異常終了を報告し(処理608)、図8に示すパスGr機能状態決定処理に移行し(処理608-1)、書き込み処理を終了する(処理609)。

【0042】以上のように処理することにより、正常なパスGrが選択されて書き込まれる。第1の実施例での書き込み処理のフローの説明を終わる。

【0043】次に、図7を参照して読み出し処理のフローを説明する。まず、マイクロプロセッサ1023は、使用するパスGrを選択する。これは、メモリ1024の領域305に記憶している全てのパスGrの機能状態を参照して読み出し専用のパスGrと正常のパスGrとを探す。候補のパスGrが複数あれば、使用頻度の低いもの、読み出し専用のもの順でパスGrを選択する(処理701)。以上の処理で読み出しに使用可能なパスGrを選択できなければCPU101に異常終了を報告し、読み出し処理を終了する(処理702)。使用するパスGrを選択できたら、そのパスGrが接続するDVAを選択する(処理703)。読み出しデータを磁気ディスク装置、DVA、BUFの順に転送する(処理704)。転送に失敗したら異常終了を報告し(処理705)、図8に示すパスGr機能状態決定処理に移行し(処理705-1)、読み出し処理を終了する。転送に成功したら、使用したパスGrで全ての磁気ディスク装置から読み出しができるかを調べる。もしできないのなら、読み出したデータに基づいて読み出せないデータをBUF1022に生成する(処理706、707)。全てのデータが揃ったところでECCによるデータの検査をし(処理708)、不合格ならCPUに異常終了を報告し、読み出し処理を終了する(処理709)。合格ならデータレコードとしてBUF、CHA、CPUの順に転送し、CPU101に正常終了を報告し、読み出し処理を終了する(処理710、711)。第1の実施例での読み出し処理のフローの説明を終わる。

【0044】以上のように処理することにより、故障し

ているパスを含むパスGrに対しては、読み出し専用として、読み出せなかったデータに対しては、読みだしたデータに基づいて生成することができる。さらに、全てのパスGr間で入出力可能なディスク装置を一致させないで、図26に示すように故障しているパスを含むパスGrに対して読み出し専用とするだけでもよい。

【0045】上記の構成をとると、転送可能な記憶装置が多いパスGrを使用して、ECC-Grに書き込みを行い、データレコードを分割し、ECCを付加してできる複数のデータの多くを記憶装置に書き込める。また、読み出し専用のパスGrで古いデータを読み出さなくなる。よって、記憶制御装置の信頼性が向上する。

【0046】また、図28に示すように、ある磁気ディスク装置に接続するパスが全て故障(閉塞)した場合には、読み出し専用に行っているパスGrを書き込みと読み出しとが行える正常な状態に戻してもよい。これは、ある磁気ディスク装置に接続するパスが全て故障しているので書き込みが行われないために、正常状態にすることができる。

【0047】以上で第1の実施例の説明を終わる。

【0048】次に、第2の実施例について説明する。本実施例の構成図は、第1の実施例と同一であり説明を省略する。第2の実施例について図29を参照して説明する。第2の実施例を説明するための模式図を図29に示す。第2の実施例の特徴は、論理閉塞扱いしていた正常のパス、つまり、全てのパスGr間で入出力可能なディスク装置を一致させているときに、論理閉塞しているパスを、データ転送後に正常状態に戻すところにある。

【0049】まず、図10を用いてメモリ1024に格納される内容を説明する。メモリ1024は書き込み処理用のプログラムの領域1001と、読み出し処理用のプログラムの領域1002と、パスGr機能状態決定プログラムの領域1003と、パスの状態の領域1004と、パスGrの機能状態の領域1005とに分かれている。パスの状態の領域1004と、パスGrの機能状態の領域1005との内容は、第1の実施例のメモリ1024の領域304と領域305のそれぞれと同一なので説明を省略する。

【0050】また、書き込み処理のフローと、読み出し処理のフローも第1の実施例と同一なので説明を省略する。

【0051】次に、図11を用いて第2の実施例のパスGr機能状態決定のフローを説明する。図11に示すフローは、前述の図10に示すメモリ1024に格納されるパスGr機能状態決定プログラムのフローを示している。

【0052】図11において、マイクロプロセッサ1023は、メモリ1024の領域1004を参照して故障または論理閉塞のパスを数える。故障または論理閉塞のパスが3個以上あるパスGrを閉塞にする(処理110

1)。処理1101で閉塞のバスGrが生じたら、閉塞でないバスGrを正常として決定し(処理1103)。閉塞のバスGrがなければ、バスGrに閉塞がない時のバスGr機能状態決定処理をおこなう(処理1104)。最後に、機能状態が正常であるバスGrを構成するバスの中に論理閉塞のものがあるか否かを判定し、論理閉塞があれば、そのバスが接続する磁気ディスク装置のデータを回復し、バスの状態を論理閉塞から正常にする(処理1105、1106)。

【0053】バスGrに閉塞がない時のバスGr機能状態決定処理1104は第1の実施例のバスGrに閉塞がない時のバスGr機能状態決定処理804と同一であり説明を省略する。

【0054】上記の構成をとると、第1の実施例と同じ効果が望める。また、記憶制御装置が読みだしたデータから記憶装置のデータを回復するので、データレコードを分割し、ECCを付加しててできる複数のデータを第1の実施例より多く記憶装置に書き込める。よって、記憶制御装置の信頼性が第1の実施例より向上する。

【0055】以上で第2の実施例の説明を終わる。

【0056】次に、第3の実施例について説明する。本実施例の構成図は、第1の実施例と同一であり説明を省略する。第3の実施例について図30を参照して説明する。第3の実施例を説明するための模式図を図30に示す。第3の実施例の特徴は、ECC-Grに対する1回の入出力を複数のバスGrを組み合わせて用いておこなうところである。すなわち、書き込み時に、すべてのバスGrのバスのうち、故障していない正常なバスをより多く含むバスGrを利用して、磁気ディスク装置に対して書き込みを行う。このとき、バスGrとして、前述のバスGr1およびバスGr2のほかに、全てのバスのなかから正常なバスのみを選択するバスGrを備えるようにできる。

【0057】まず、図12を参照してメモリ1024の内容を説明する。メモリ1024は、書き込み処理用のプログラムの領域1201と、読み出し処理用のプログラムの領域1202と、組合せバスGr機能状態決定プログラムの領域1203と、バスの状態の領域1204と、組合せバスGrの機能状態の領域1205とに分かれている。ここで、組合せバスGrとは、バスGr1と、バスGr2と、すべてのバスGrのバスのうち、磁気ディスク装置に対応し、故障していない正常なバスのみを選択して利用するバスGr(選択転送手段、以下、選択バスGrという)とをいう。本実施例においては、これらの組合せバスGrのなかからより正常なバスが多いバスGrを利用して読み出しと書き込みとを行う。

【0058】バスの状態の領域1204の内容は、第1の実施例のメモリ1024の領域304の内容と同一なので説明を省略する。

【0059】図13を参照して組合せバスGrの機能状

態の領域1205を説明する。領域1205は、バスGr1026-1の機能状態を表す領域1301と、バスGr1026-2の機能状態を表す領域1302と、選択バスGrの機能状態を表す領域1303とに分かれている。バスGrの機能状態そのものは、第1の実施例と同一なので説明を省略する。

【0060】次に、図16を参照して組合せバスGr機能状態決定のフローを説明する。図16において、マイクロプロセッサ1023は、磁気ディスク装置に対して全く入出力できないバスの状態が、組合せバスGr間で一致するか否かを調べる(処理1601)。ここで、第3の実施例では、バスGrが1026-1と1026-2との2個のみである。よって、実際には、図4に示す、メモリ1024の領域401-1と領域402-1とに格納されているバス状態の内容が一致するか否かを調べ、同様に、領域401-2と領域402-2とに格納されているバス状態の内容、・・・、領域401-nと領域402-nとに格納されているバス状態の内容それぞれが一致するかを調べる。これにより、3つの組合せバスGrについて調べることになる。全てのバス状態が正常で一致すれば、3個の組合せバスGrを正常にする(処理1602)。一致しないならば、全く入出力できない閉塞状態のバスが少ないバスGrを正常にし、残りのバスGrを読み出し専用にする(処理1603)。第3の実施例では、バスGrが1026-1と1026-2の2個のみなので、実際には、必ず2個のバスGrを組合せたなかから選択した選択バスGrが、全く入出力できないバスが少ないバスGrになる。例えば、図30に示すように、右側のバスGrをバスGr1とし、左側のバスGrをバスGr2とし、全てのバスから正常状態のバスであって、磁気ディスク装置に対応するバスを選択したものを選択バスGrとする。この例においては、バスGr2のバス5が閉塞状態であるので、バス5とバス11とのバス状態が一致しない。この場合には、図13に示す組合せバスGrの機能状態において、選択バスGrを正常状態にする。バスGr1とバスGr2とは読み出し専用にする。もしくは、バスGr2については、全てのバスが正常であるので正常状態としてもよい。このようにして組合せバスGrの機能状態を設定する。また、選択バスGrを正常としたときに、どのバスを選択するかを記憶しておく。そして最後に、3個以上の磁気ディスク装置に対して全く入出力できないバスGrを閉塞にする(処理1604)。

【0061】バスGrの機能状態の決定処理は、記憶制御装置を初期作動させる時と、DVAがバスの故障をマイクロプロセッサ1023に報告した時とおこなう。

【0062】次に、図14を参照して書き込み処理のフローを説明する。図14において、マイクロプロセッサ1023は、書き込むデータレコードをCHA1021からBUF1022へ転送する(処理1401)。次

に、使用するパスGrを選択する。この処理は、メモリ1024の領域1205に記憶している全てのパスGrの機能状態を参照し、正常なパスGrを探す。正常状態のパスGrが複数あれば、単一のパスGr、使用頻度の低いもの、の順にパスGrを選択する。正常のパスGrがなければ、CPU101に異常終了を報告し、書き込み処理を終了する（処理1402、1403）。正常なパスGrを選択できれば、BUF1022に転送された書き込むデータレコードを（n-2）等分する（処理1404）。次に、（n-2）等分した結果のデータの集まりに対してECCを作成し、データをn個にする（処理1405）。これらn個のデータは、それぞれの磁気ディスク装置に書き込まれる。そして、このECCは、n個の磁気ディスク装置の内、磁気ディスク装置の故障によって、2個までの磁気ディスク装置から読み出しができないときに、読み出せないデータを生成できる符号である。ECCを作成した後に、選んだパスGrのパスが接続しているDVAを選択する（処理1407）。この場合において、前述の選択パスGrを選んだときに、記憶しているパスを用いて転送を行う。もしくは、選択パスGrを選んだときには、再度、図4に示すパスの状態領域を参照して正常なパスのみを利用するようにしてよい。そして、n個のデータをBUF、DVA、磁気ディスク装置の順に転送する（処理1408）。転送に失敗したらCPU101に異常終了を報告をし（処理1409）、図16に示す組合せパスGrの機能状態決定フローに移行し（処理1409-1）、書き込み処理を終了する。この後、処理1402において選んだパスGrの中に書き込みがされていない未使用のパスがあるのならば、処理1407から繰り返す（処理1406）。反復を終了した後に、CPU101に正常終了を報告する（処理1410）。以上で第3の実施例での書き込み処理のフローの説明を終わる。

【0063】次に、図15を参照して読み出し処理のフローを説明する。図15において、マイクロプロセッサ1023は、使用するパスGrを選択する。メモリ1024の領域1205に記憶している全ての組合せパスGrの機能状態を参照して正常のパスGrと読み出し専用のパスGrを探す。候補のパスGrが複数あれば、単一のパスGr、使用頻度の低いもの、読み出し専用のもの、の順に採用する。以上の処理で読み出しに使用可能なパスGrを選択できなければ、CPU101に異常終了を報告し、読み出し処理を終了する（処理1501、1502）。読み出しに使用可能なパスGrがあれば、そのパスGrが接続するDVAを選択する（処理1504）。n個のデータを磁気ディスク装置、DVA、BUFの順に転送する（処理1505）。転送に失敗したらCPU101に異常終了を報告し（処理1506）、図16に示す組合せパスGrの機能状態決定フローに移行する（処理1506-1）。処理1501において、選

んだパスGrの中に読出しされていない未使用のパスがあるのならば、処理1504から繰り返す（処理1503）。反復を終了した後に使用したパスGrで全ての磁気ディスク装置から読み出しができるかを調べる。もしできないのなら、読み出せたデータから読み出せないデータをBUF1022に生成する（処理1508、1509）。全てのデータが揃ったところでECCによるデータの検査をし（処理1510）、不合格ならCPU101に異常終了を報告し、読み出し処理を終了する（処理1511）。合格ならデータレコードとしてBUF、CHA、CPUの順に転送し、CPU101に正常終了を報告し、読み出し処理を終了する（処理1512、1513）。

【0064】以上で第3の実施例での読み出し処理のフローの説明を終わる。

【0065】上記の処理を行うと、第1の実施例と同じ効果が望める。また、少なくとも一方のパスから入出力できる磁気ディスク装置は入出力できることになるので、データレコードを分割し、ECCを付加してできる複数のデータを第1や第2の実施例より多く記憶装置に書き込める。よって、記憶制御装置としての信頼性が第1や第2の実施例より向上する。

【0066】以上で第3の実施例の説明を終わる。

【0067】次に、第4の実施例について説明する。本実施例の構成図は、第1の実施例と同一であり説明を省略する。第4の実施例では、図31に模式図を示すように、ディスク装置に対して読み書きを行う際に、パスが使用不可能な状態を検知し、不可能な場合には、そのパスに接続しているディスク装置に読み書きを行わないようにするために該ディスク装置に接続している他のパスに対して論理閉塞を行うように制御する場合について説明する。すなわち、第4の実施例の特徴は、パスに故障があれば、そのパスが接続する記憶装置に接続する他のパスを論理的に閉塞にすることである。

【0068】図17を参照してメモリ1024に格納される内容を説明する。メモリ1024は、書き込み処理用のプログラムの領域1701と、読み出し処理用のプログラムの領域1702と、論理閉塞パス設定プログラムの領域1703と、パスの状態の領域1704とに分かれている。パスの状態の領域1704は、第1の実施例のメモリ1024の領域304の内容と同一である。

【0069】次に、論理閉塞パス設定プログラムの領域1703に格納されている論理閉塞パス設定プログラムの処理フローについて図20を参照して説明する。図20において、マイクロプロセッサ1023は、メモリ1024の領域1704に格納されるパスの状態を参照して、個々の記憶装置に接続するパスが故障していないかを調べる。故障しているパスがあれば、その記憶装置に接続する他の故障していないパスを論理閉塞（論理閉塞）に設定する（処理2002、2003）。以上の処

理をECC-Grを構成する記憶装置全てに対しておこなう。バスGrの機能状態の決定処理は、記憶制御装置を初期作動させる時と、DVAがバスの故障をマイクロプロセッサ1023に報告した時とにおこなう。以上で第4の実施例の論理閉塞バス設定処理のフローの説明を終わる。

【0070】つぎに、図18を参照して第4の実施例の書き込み処理のフローを説明する。図18において、マイクロプロセッサ1023は、書き込むデータレコードをCHA1021からBUF1022へ転送する(処理1801)。次に、BUF1022のデータレコードを(n-2)等分する(処理1802)。この(n-2)個のデータの集まりに対してECCを作成し、データをn個にする(処理1803)。これらn個のデータはそれぞれ1個の磁気ディスク装置に書き込まれる。そして、このECCは、n個の磁気ディスク装置の内、磁気ディスク装置の故障によって、任意の2個までの磁気ディスク装置から読み出しができないときに、読み出せないデータを生成できる符号である。ECCを作成した後に、2個のバスGrのうち一方を選び、選んだバスGrが接続するDVAを選択する(処理1804)。そして、n個のデータをBUF、DVA、磁気ディスク装置の順に転送する(処理1805)。転送に成功したらCPU101に正常終了を報告し、失敗したらCPU101に異常終了を報告し、書き込み処理を終了する(処理1806、1807)。以上で第4の実施例での書き込み処理のフローの説明を終わる。

【0071】次に、図19を参照して第4の実施例の読み出し処理のフローを説明する。まず、2個のバスGrのうち一方を選び、選んだバスGrが接続するDVAを選択する(処理1901)。読み出しデータを磁気ディスク装置、DVA、BUFの順に転送する(処理1902)。転送に失敗したらCPU101に異常終了を報告し、読み出し処理を終了する(処理1903)。転送に成功したら、使用したバスGrで全ての磁気ディスク装置から読み出しができるか否かを調べる。もしできないのなら、読み出せたデータから読み出せないデータをBUF1022に生成する(処理1904、1905)。全てのデータが揃ったところでECCによるデータの検査をし(処理1906)、不合格ならCPU101に異常終了を報告し、読み出し処理を終了する(処理1907)。合格ならデータレコードとしてBUF、CHA、CPUの順に転送し、CPU101に正常終了を報告し、読み出し処理を終了する(処理1908、1909)。以上で第4の実施例での読み出し処理のフローの説明を終わる。

【0072】上記の構成とすると、入出力可能な記憶装置が同一のECC-Grに接続するバスGr間で一致する。つまり、書き込み時に使用したバスGrか否かにかかわらずデータレコードとしての読み出しができる。よ

って、記憶制御装置の信頼性が向上する。以上で第4の実施例の説明を終わる。

【0073】次に、第5の実施例について説明する。本実施例の構成図は、第1の実施例と同一であり説明を省略する。第5の実施例の特徴は、バスに故障があるときECC-Grに対する1回の入出力を全てのバスGrを参照しておこなうところである。すなわち、はじめに、バスに故障があるか否かを調べて、故障がある場合には各バスGrを介して読み書きの処理をそれぞれ行う。これにより、読み書き動作は、バスGrの数だけ行われるが、故障しているバスがあっても他のバスGrのバスから書き込みが行うことができる。

【0074】まず、図21を参照してメモリ1024に格納される内容を説明する。メモリ1024は書き込み処理用のプログラムの領域2101と、読み出し処理用のプログラムの領域2102と、バスの状態の領域2103と、に分かれている。バスの状態の領域2103の内容は第1の実施例のメモリ1024の領域304の内容と同一である。

【0075】次に、図22を参照して第5の実施例の書き込み処理のフローを説明する。図22において、マイクロプロセッサ1023は、書き込むデータレコードをCHA1021からBUF1022へ転送する(処理2201)。次に、BUF1022のデータを(n-2)等分する(処理2202)。この(n-2)個のデータの集まりに対してECCを作成し、データをn個にする(処理2203)。これらn個のデータはそれぞれの磁気ディスク装置に書き込まれる。そして、このECCは、n個の磁気ディスク装置の内、磁気ディスク装置の故障によって、任意の2個までの磁気ディスク装置から読み出しができないときに、読み出せないデータを生成できる符号である。ECCを作成した後に、メモリ1024の領域2103バスの状態を参照して、バスに故障があるかを調べる(処理2204)。故障があれば、バスに故障があるときの書き込み処理をおこなう(処理2205)。故障がなければ、2個のバスGrのうち一方を選び、選んだバスGrが接続するDVAを選択する(処理2206)。そして、n個のデータをBUF、DVA、磁気ディスク装置の順に転送する(処理2207)。転送に成功したらCPU101に正常終了を報告し、失敗したらCPU101に異常終了を報告し、書き込み処理を終了する(処理2208、2209)。

【0076】つぎに、図23を参照し、バスに故障があるときの書き込み処理のフローを説明する。図23において、任意のバスGrを選び、そのバスGrが接続するDVAを選択する。そして、n個のデータをBUF、DVA、磁気ディスク装置の順に転送する。転送に失敗したらCPU101に異常終了を報告して処理を終わる。

(処理2302、2303、2304)。以上の処理を書き込み対象のECC-Grに接続するバスGrの全て

に対しておこなう。すなわち、順次バスGrを選び、それぞれ転送処理を行う。以上で第5の実施例での書き込み処理のフローの説明を終わる。

【0077】次に、図24を参照して第5の実施例の読み出し処理のフローを説明する。図24において、マイクロプロセッサ1023は、メモリ1024の領域2103のバスの状態を参照して、バスに故障があるかを調べる(処理2401)。故障があれば、バスに故障があるときの読み出し処理をおこなう(処理2402)。故障がなければ、2個のバスGrのうち一方を選び、選んだバスGrが接続するDVAを選択する(処理2403)。読み出しデータを磁気ディスク装置、DVA、BUFの順に転送する(処理2404)。転送に失敗したらCPU101に異常終了を報告し、読み出し処理を終了する(処理2405)。転送に成功したら、使用したバスGrで全ての磁気ディスク装置から読み出しができるかを調べる。もしできないのなら、読み出せたデータから読み出せないデータをBUF1022に生成する(処理2406、2407)。全てのデータが揃ったところでECCによるデータの検査をし(処理2408)、不合格ならCPU101に異常終了を報告し、読み出し処理を終了する(処理2409)。合格ならデータレコードとしてBUF、CHA、CPUの順に転送し、CPU101に正常終了を報告し、読み出し処理を終了する(処理2410、2411)。

【0078】図25を参照してバスに故障があるときの読み出し処理のフローを説明する。任意のバスGrを選び、そのバスGrが接続するDVAを選択する。そして、n個のデータをDVA、磁気ディスク装置、BUFの順に転送する。転送に失敗したらCPU101に異常終了を報告して処理を終わる。(処理2502、2503、2504)。以上の処理を読み出し対象のECC-Grに接続するバスGrの全てに対しておこなう。すなわち、順次バスGrを選び、それぞれ転送処理を行う。以上で第5の実施例での読み出し処理のフローの説明を終わる。

【0079】上記の構成をとると、どれか1個のバスGrでも入出力できる記憶装置であればデータレコードを分割して、ECCを作成した結果できる複数のデータの書き込み読み出しができる。よって、データの冗長度を高く保て、記憶制御装置の信頼性が向上する。以上で第5の実施例の説明を終わる。

【0080】以上説明したように、上記5つの実施例によれば、バスの故障の少ないバスGrを書き込みと読み出しとに使用可能とする。残りのバスGrの中で、故障したバスがECCで訂正可能な個数以内のものを読み出しのみに使用可能とし、その他のバスGrを書き込みと読み出しに使用不可能とする。そして、書き込みと読み出しに使用可能なバスGrが入出力できない記憶装置に接続する全てのバスを故障扱いにする。これらより、E

CC-Grに対する1回の入出力を1個のバスGrでおこなう方式を取るときに、データレコードを分割して、ECCを付加してできる複数のデータを記憶している記憶装置の個数を本発明を採用しない場合と比べ増加できる。また、読み出しのみに使用可能なバスGrで古いデータを読み出ししてしまうことを防ぐ。結果、記憶制御装置の信頼性が向上する。

【0081】また、上記実施例では、書き込みと読み出しとに使用可能なバスGrの中に故障扱いを受けているバスがあれば、そのバスの接続する記憶装置のデータを回復し、故障扱いをとりやめる。この結果、人員以外による記憶装置のECC-Grへの再組み込がおこるので、記憶制御装置の信頼性が向上する。

【0082】また、別の手段として、記憶装置に接続するバスの内、1個でも故障したらその記憶装置に接続するバスを全て論理的に閉塞とする。この結果、データレコードを分割して、ECCを作成した結果できる複数のデータを書き込める記憶装置が同じECC-Grに接続するバスGr間で一致する。よって、書き込み時に使用したバスGrが否かにかかわらずデータレコードとしての読み出しができる。

【0083】さらに、別の手段として、ECC-Grに対する書き込みと読み出しに使用可能か否かを表す機能状態をバスGrの組合せ毎に決定記憶しておき、ECC-Grに対して書き込みまたは読み出しの要求を処理するバスGrを決める。また、別の手段として、各バスGrで入出力できる記憶装置が一致しない時、組合せに必ず全てのバスGrを用いる。これらより、ECC-Grに対する1回の入出力を複数のバスGrでおこなうので、記憶装置に接続するバスの内、1個でも使用可能ならば、その記憶装置に書き込みができる。よって、正しいデータを記憶している記憶装置の個数を常にその時点での最大になる。

【0084】書き込み不可の制限を付けたバスグループを設けることで、記憶装置のデータ破壊を最小にできる。また、論理閉塞のバスを設けることで、データの訂正不可能の回数を減らす。よって、記憶制御装置の信頼性の向上を望める。

【0085】

【発明の効果】本発明によれば、記憶制御装置の信頼性が向上する。

【図面の簡単な説明】

【図1】本発明の一実施例を示す構成図。

【図2】解決すべき課題を示す説明図。

【図3】第1の実施例のメモリの内容を示す説明図。

【図4】バス状態の領域の内訳を示す説明図。

【図5】バスGrの機能状態の領域の内訳を示す説明図。

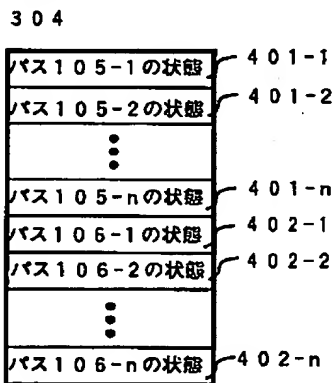
【図6】第1の実施例の書き込み処理フロー。

【図7】第1の実施例の読み出し処理フロー。

- 【図8】第1の実施例のバスGr機能状態決定フロー。
 【図9】閉塞のバスGrがない時のバスGr機能状態決定フロー。
 【図10】第2の実施例のメモリの内容を示す説明図。
 【図11】第2の実施例のバスGr機能状態決定フロー。
 【図12】第3の実施例のメモリの内容を示す説明図。
 【図13】組合せバスGrの機能状態の領域の内訳を示す説明図。
 【図14】第3の実施例の書き込み処理フロー。
 【図15】第3の実施例の読み出し処理フロー。
 【図16】組合せバスGr機能状態決定フロー。
 【図17】第4の実施例のメモリの内容を示す説明図。
 【図18】第4の実施例の書き込み処理フロー。
 【図19】第4の実施例の読み出し処理フロー。
 【図20】第4の実施例の論理閉塞バス設定処理フロー。
 【図21】第5の実施例のメモリの内容を示す説明図。
 【図22】第5の実施例の書き込み処理フロー。
 【図23】バスに故障があるときの書き込み処理フロー。
 【図24】第5の実施例の読み出し処理フロー。
 【図25】バスに故障があるときの読み出し処理フロー。
 【図26】記憶制御装置における入出力制御（読み出し専用）を説明するための模式図。

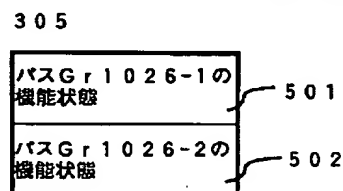
【図4】

図4 バスの状態の領域の内訳



【図5】

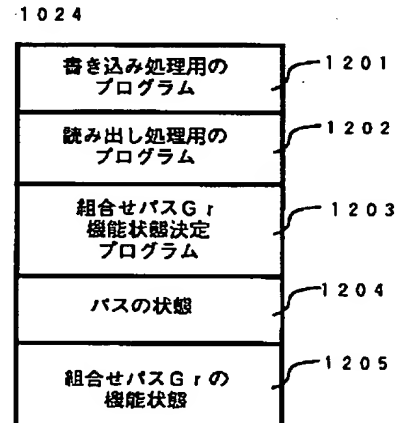
図5 バスGrの機能状態の領域の内訳



- 【図27】記憶制御装置における入出力制御（読み出し専用／論理閉塞）を説明するための模式図。
 【図28】記憶制御装置における入出力制御（両方のバスが閉塞の場合）を説明するための模式図。
 【図29】記憶制御装置における入出力制御（自動回復）を説明するための模式図。
 【図30】記憶制御装置における入出力制御（補完書き込み）を説明するための模式図。
 【図31】記憶制御装置における入出力制御（論理閉塞）を説明するための模式図。
 【図32】情報処理システムのハードウェア構成図。
 【符号の説明】
 101…中央処理装置、102…記憶制御装置、1021…チャンネルアダプタ、1022…バッファ、1023…マイクロプロセッサ、1024…メモリ、1025-1・1025-2…デバイスアダプタ、105-1・105-2…および105-n…デバイスアダプタ1025-1と磁気ディスク装置との転送路、106-1・106-2…および106-n…デバイスアダプタ1025-2と磁気ディスク装置との転送路、103-1・103-2…および103-n…磁気ディスク装置、104…ECCグループ、1026-1および1026-2…バスグループ、a01-1・a01-2…a01-n…磁気ディスク装置とのインターフェース。

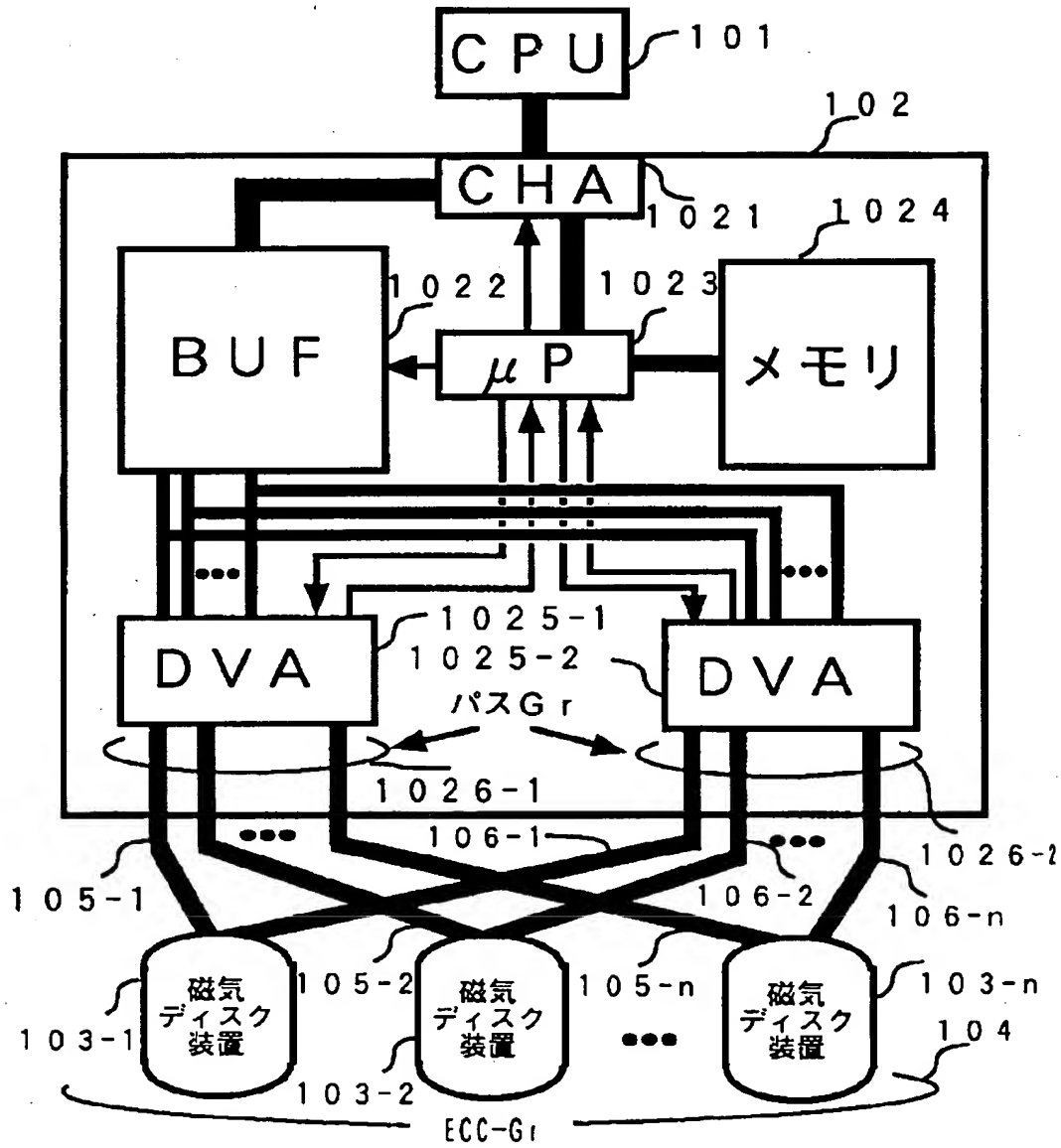
【図12】

図12 第3の実施例のメモリの内容



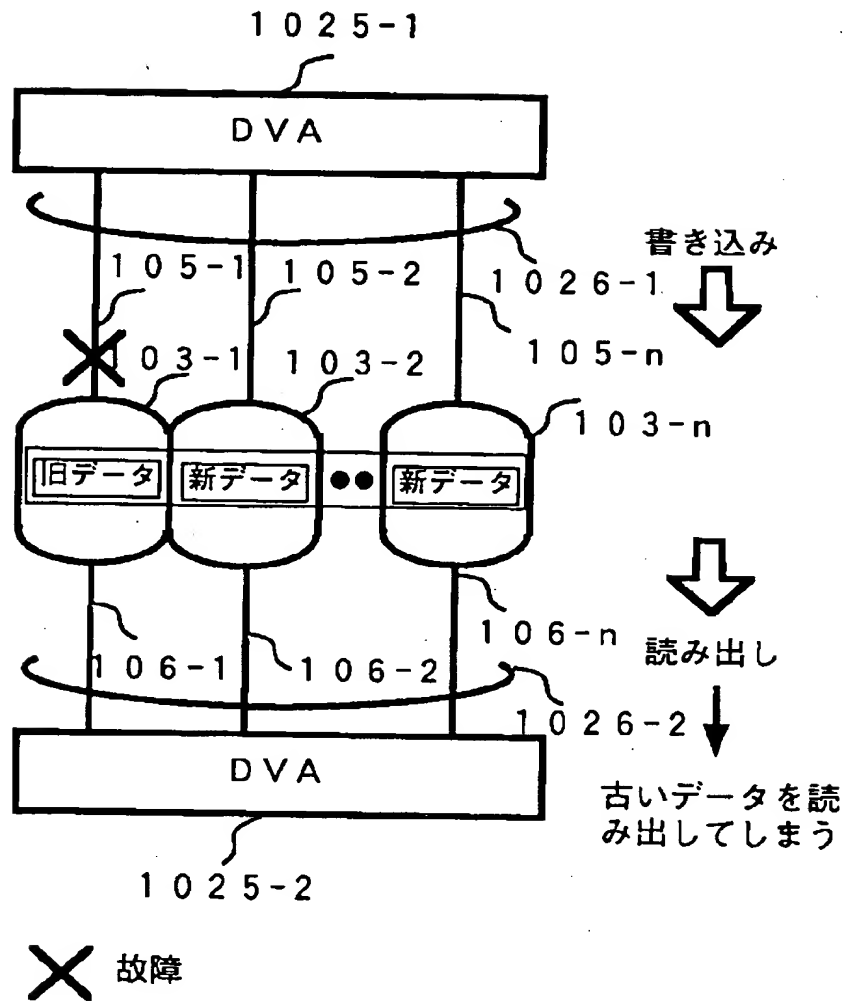
【図1】

図1 本発明の一実施例



【図2】

図2 解決すべき課題



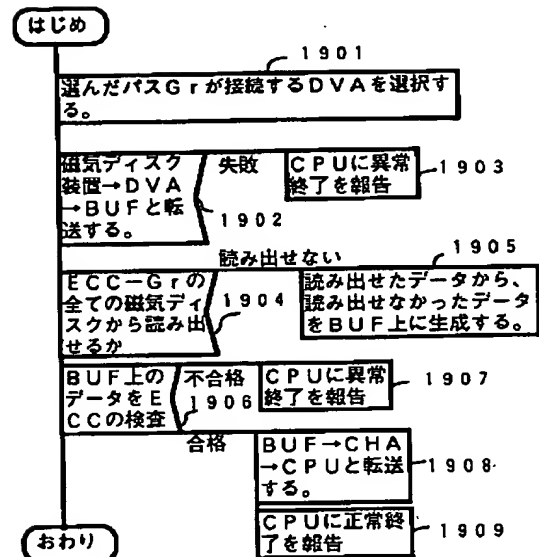
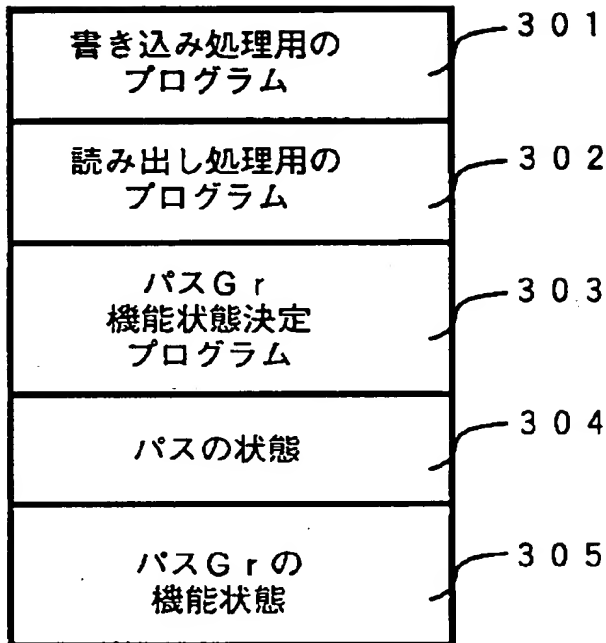
【図3】

【図19】

図3 第1の実施例のメモリの内容

図19 第4の実施例の読み出し処理フロー

1024



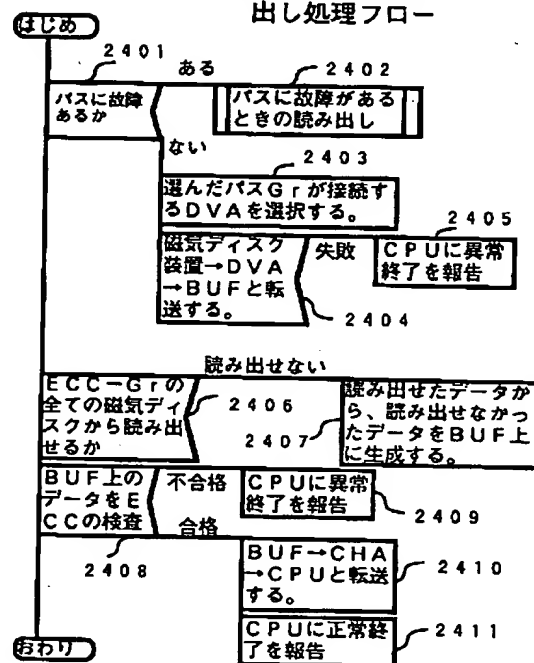
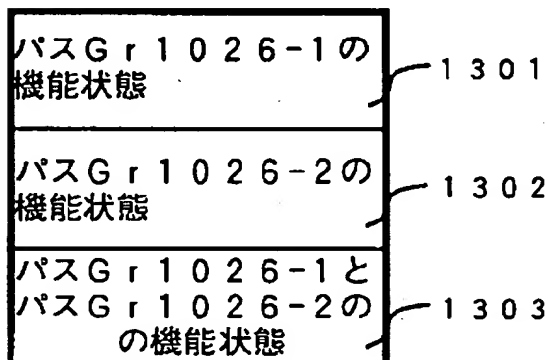
【図13】

【図24】

図13 組合せパスGrの機能状態の領域の内訳

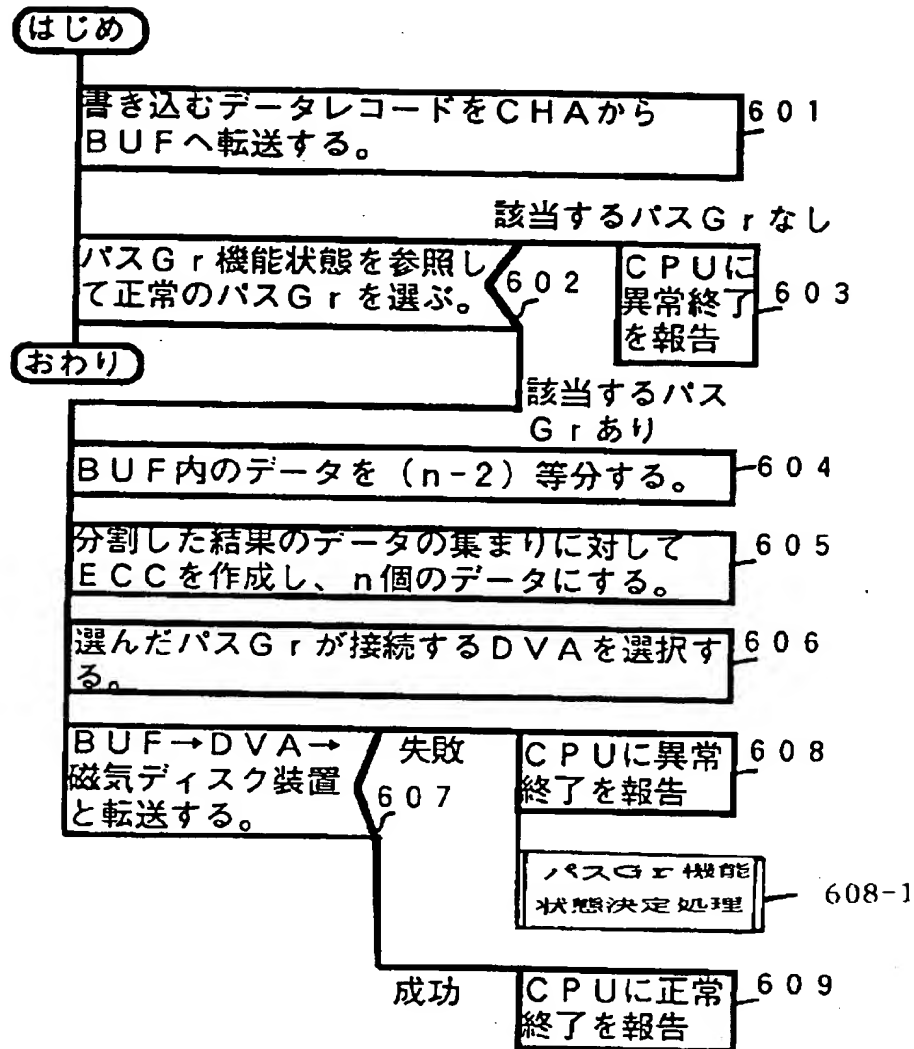
図24 第5の実施例の読み出し処理フロー

1205



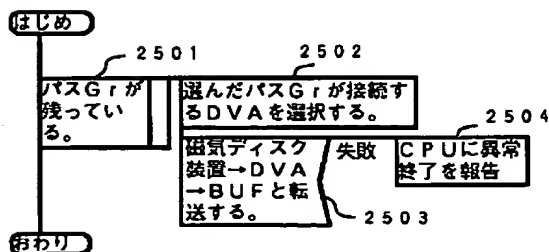
【図6】

図6 第1の実施例の書き込み処理フロー



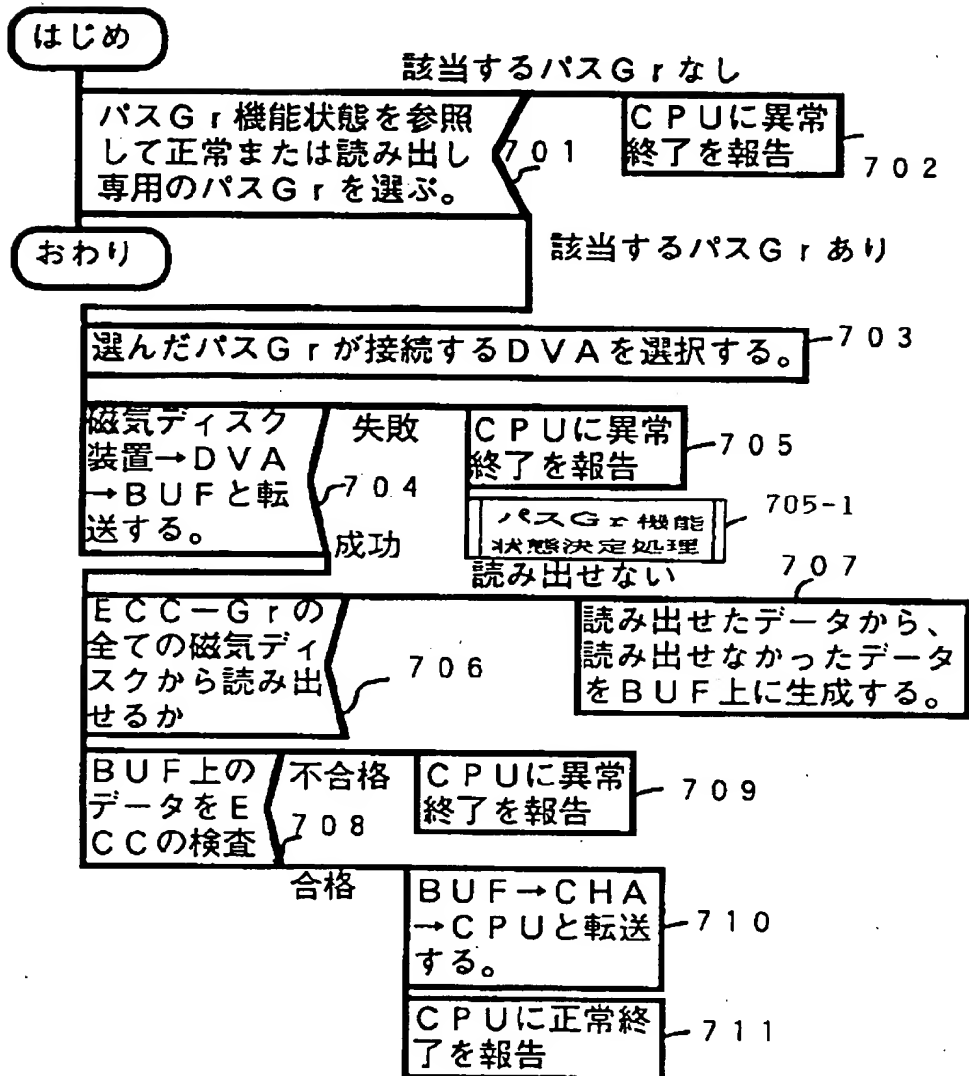
【図25】

図25 バスに故障があるときの読み出し処理フロー



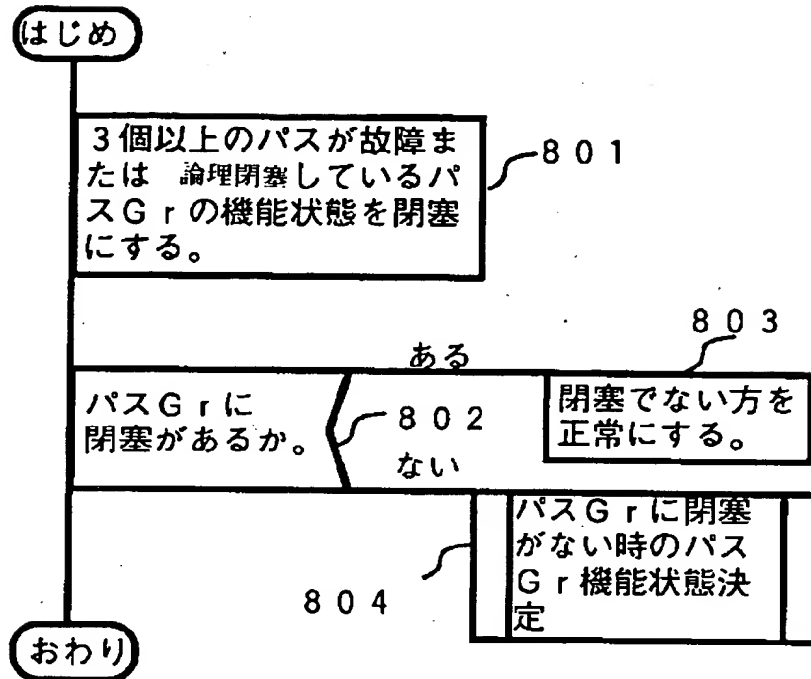
【図7】

図7 第1の実施例の読み出し処理フロー



【図8】

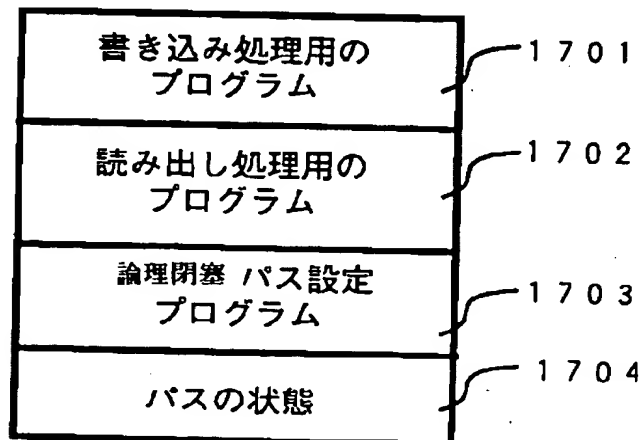
図8 第1の実施例のバスGr機能状態決定フロー



【図17】

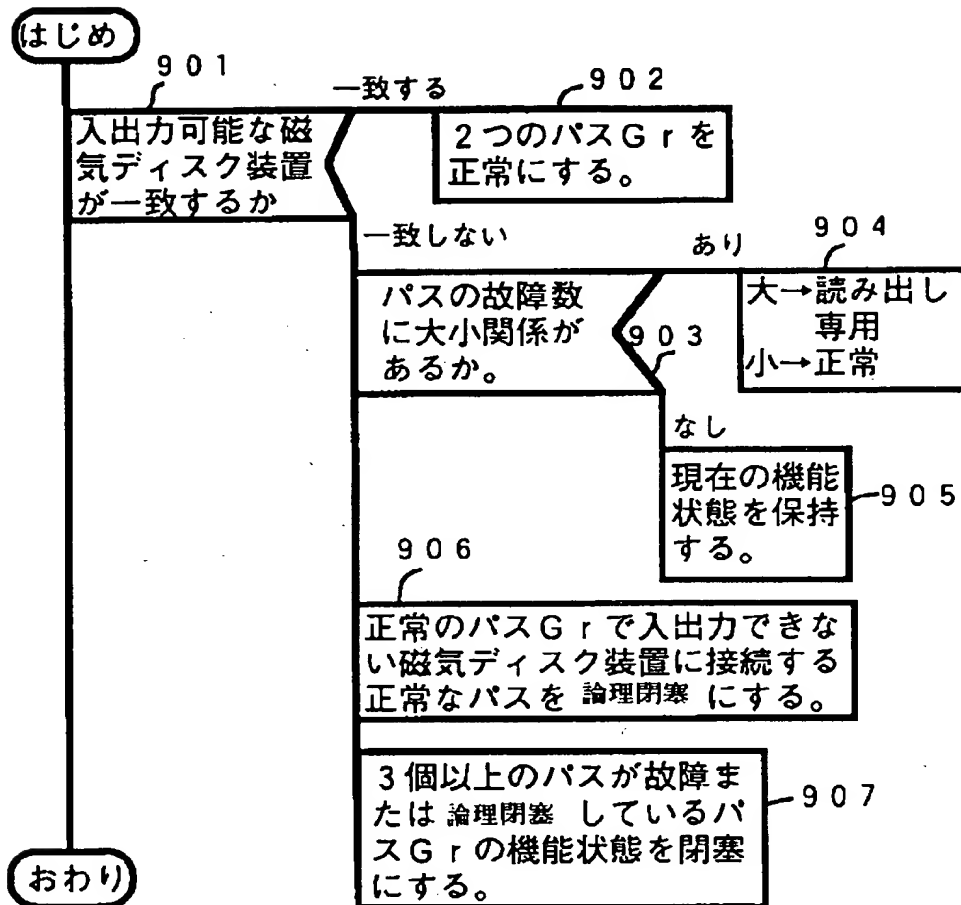
図17 第4の実施例のメモリの内容

1024



【図9】

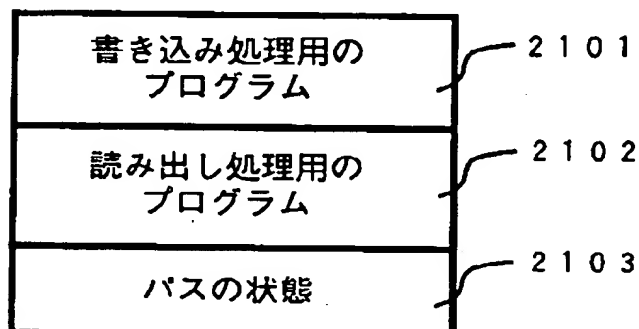
図9 閉塞のバスGrがない時のバスGr機能状態決定フロー



【図21】

図21 第5の実施例のメモリの内容

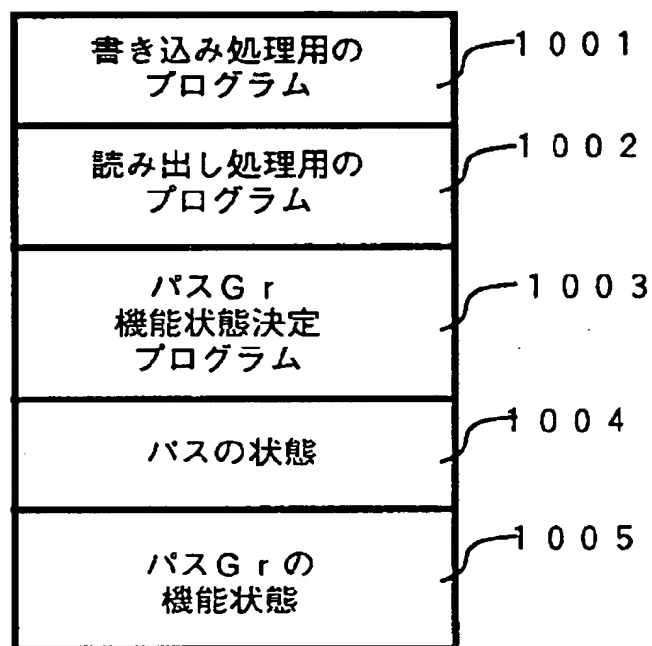
1024



【図10】

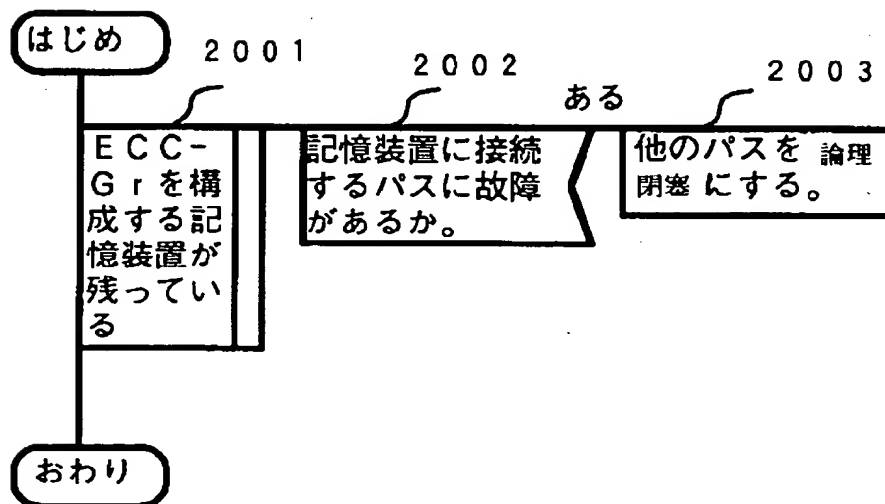
図10 第2の実施例のメモリの内容

1024



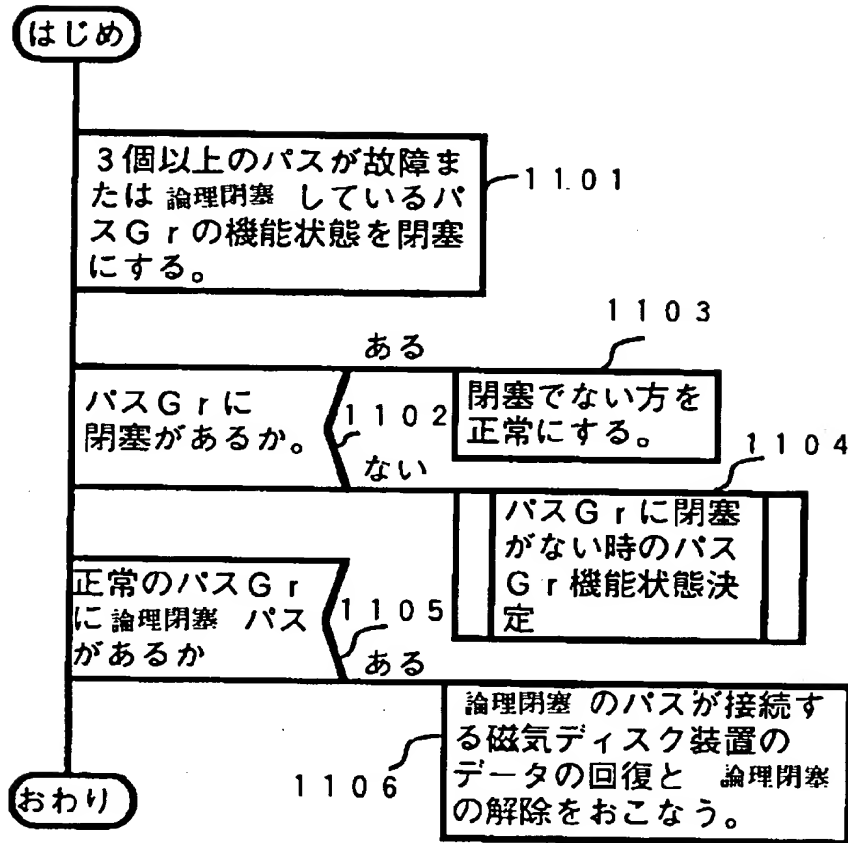
【図20】

図20 第4の実施例の擬似故障パス設定処理フロー



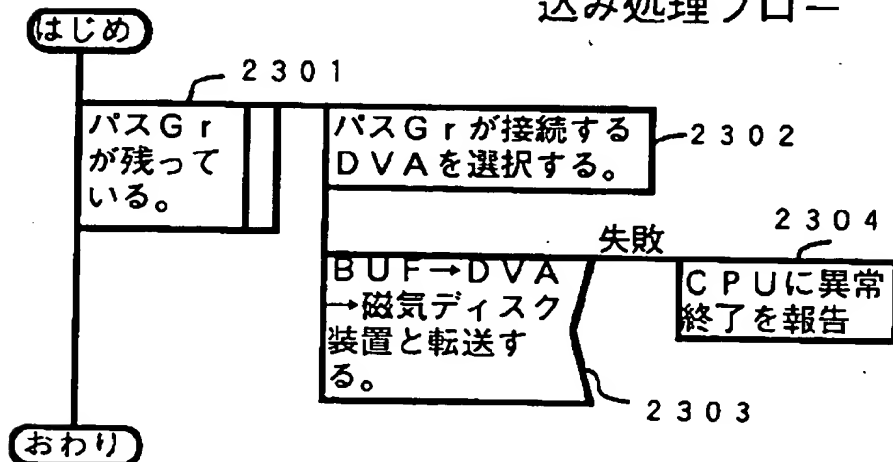
【図11】

図11 第2の実施例のバスGr機能状態決定フロー



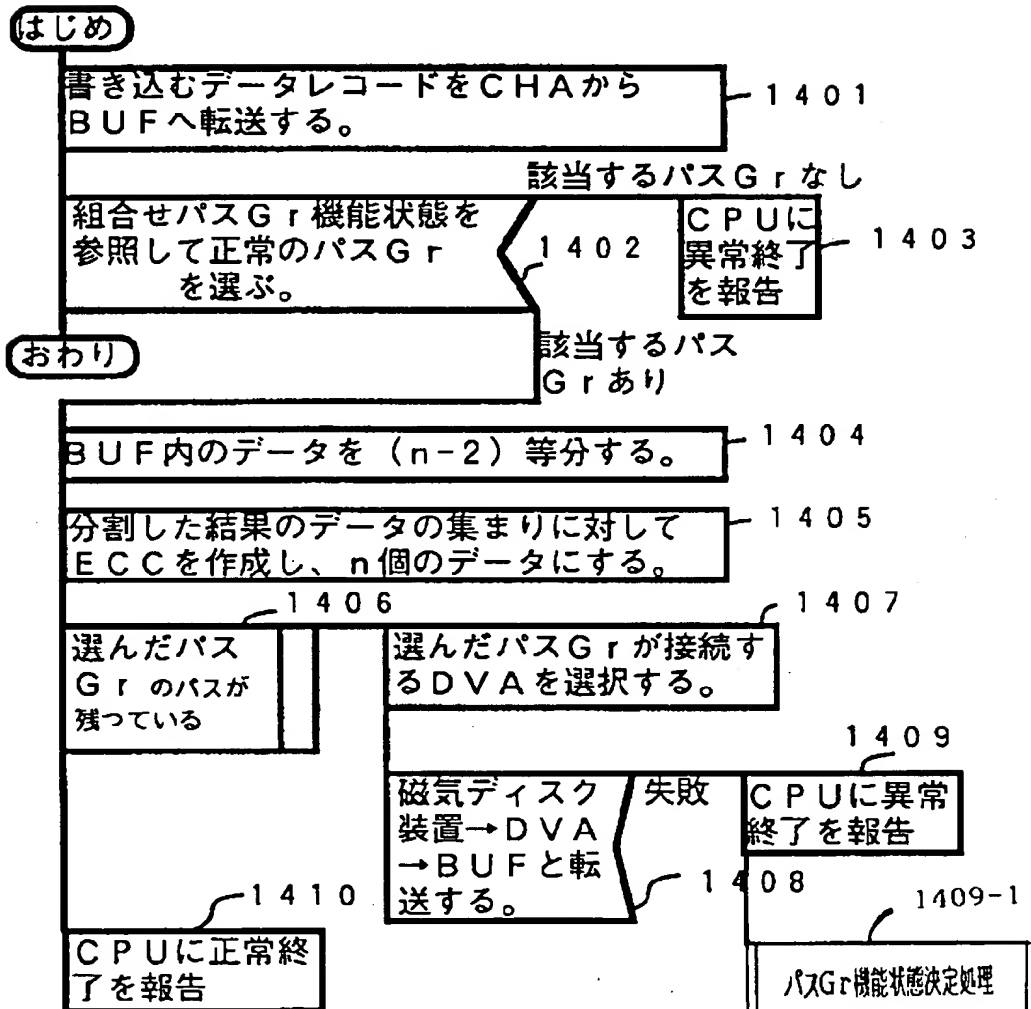
【図23】

図23 バスに故障があるときの書き込み処理フロー



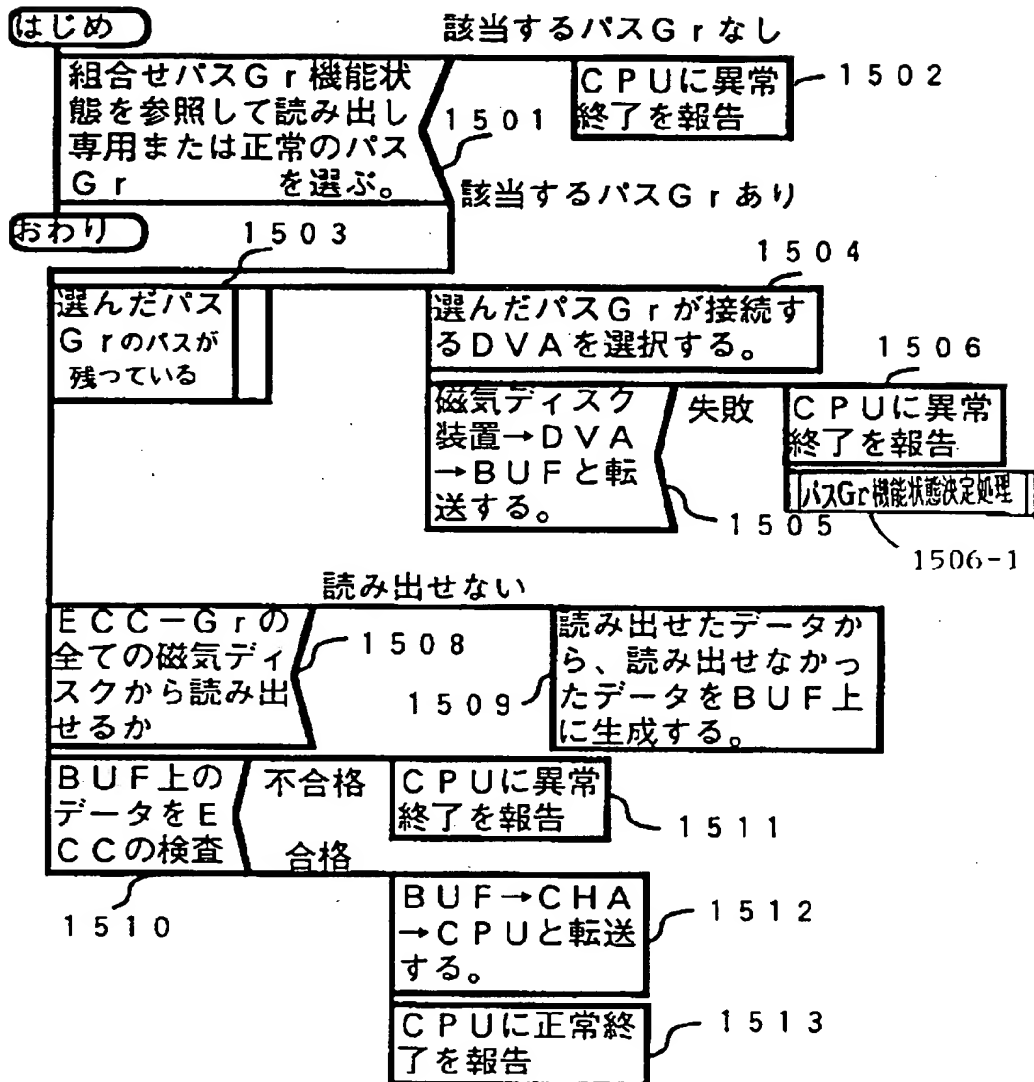
【図14】

図14 第3の実施例の書き込み 処理フロー

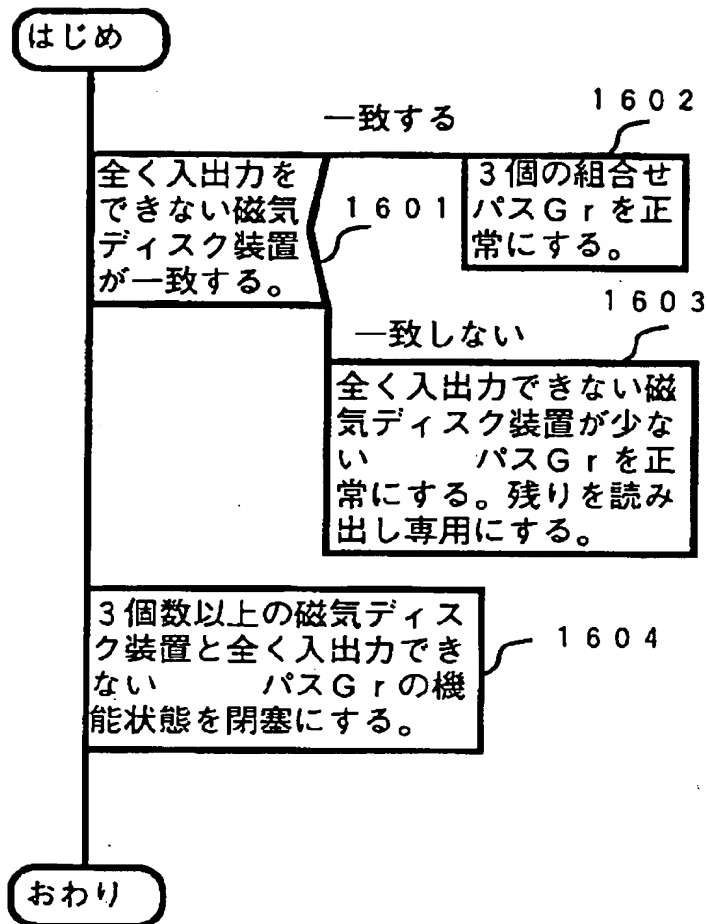


【図15】

図15 第3の実施例の読み出し処理フロー



【図16】

図16 組合せパスGr機能
状態決定フロー

【図30】

図30 補完書き方式

両方のRAIDパスを用いて書き込む

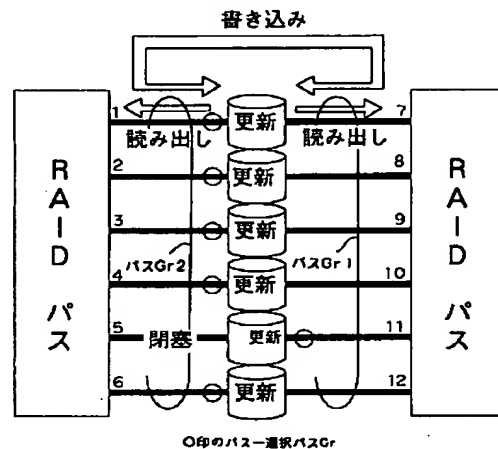


図 18 第 4 の実施例の書き込み処理フロー

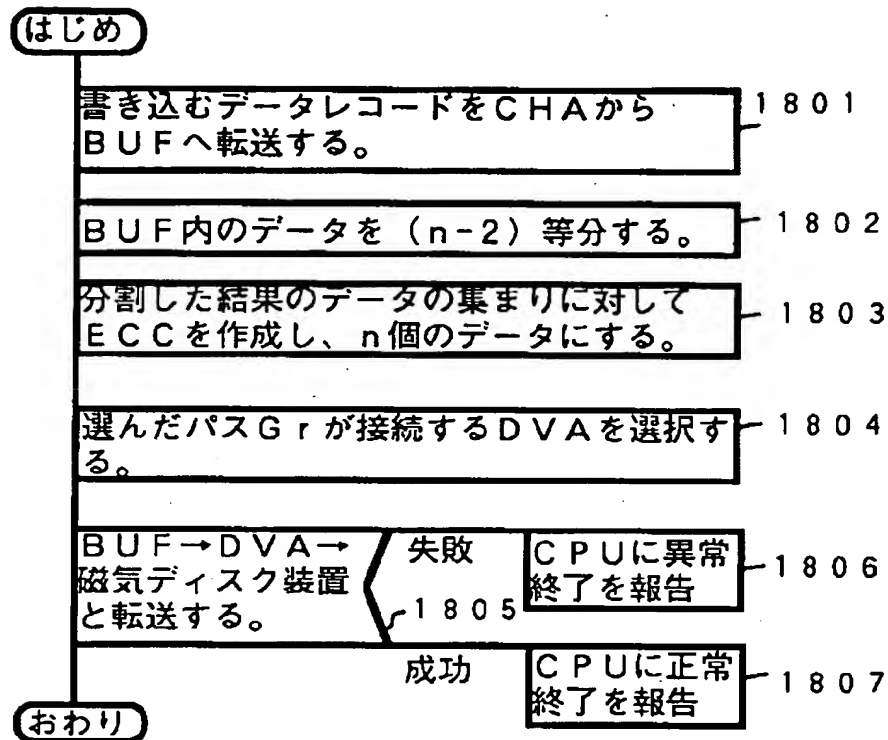
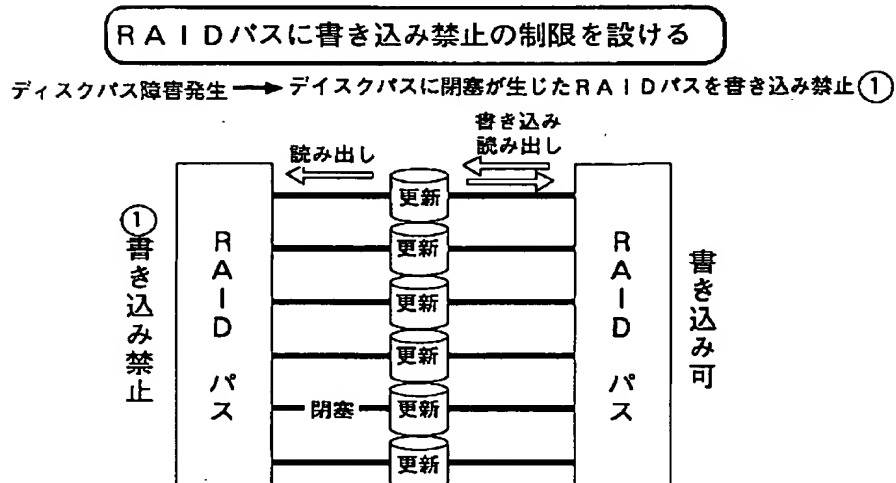


図26 入出力制限方式 読み出し専用



【図22】

図22 第5の実施例の書き込み処理フロー

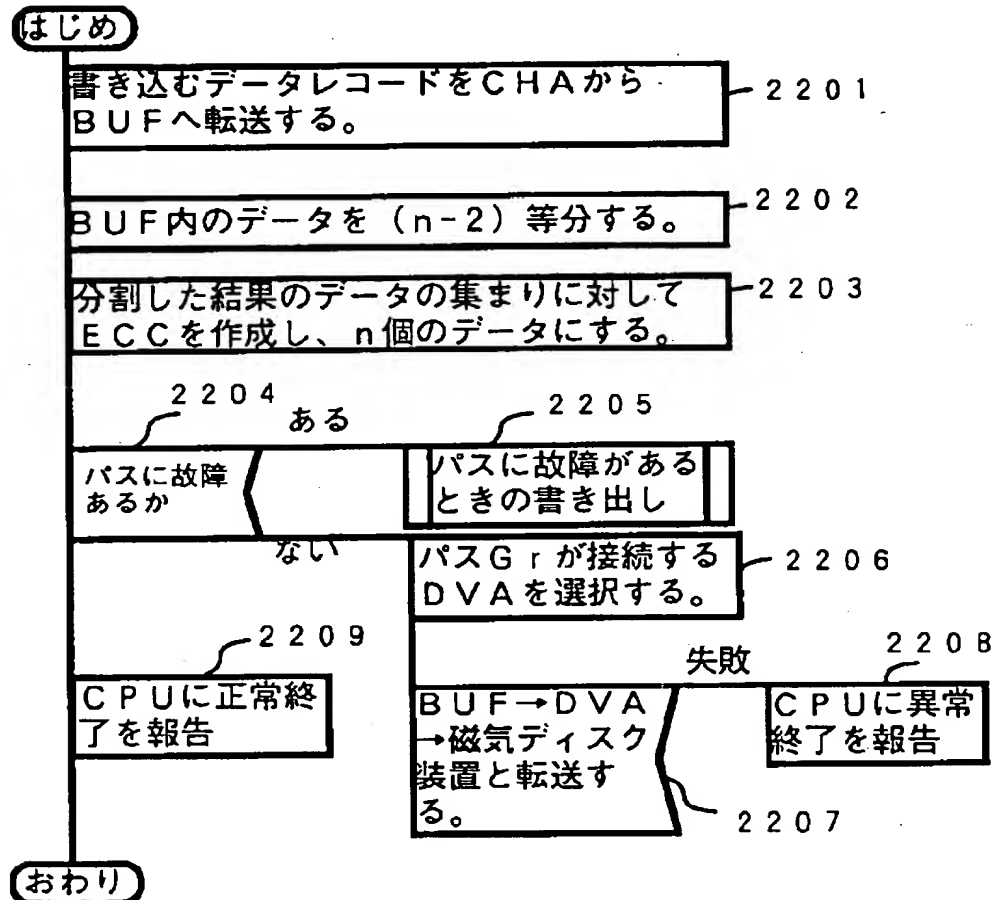
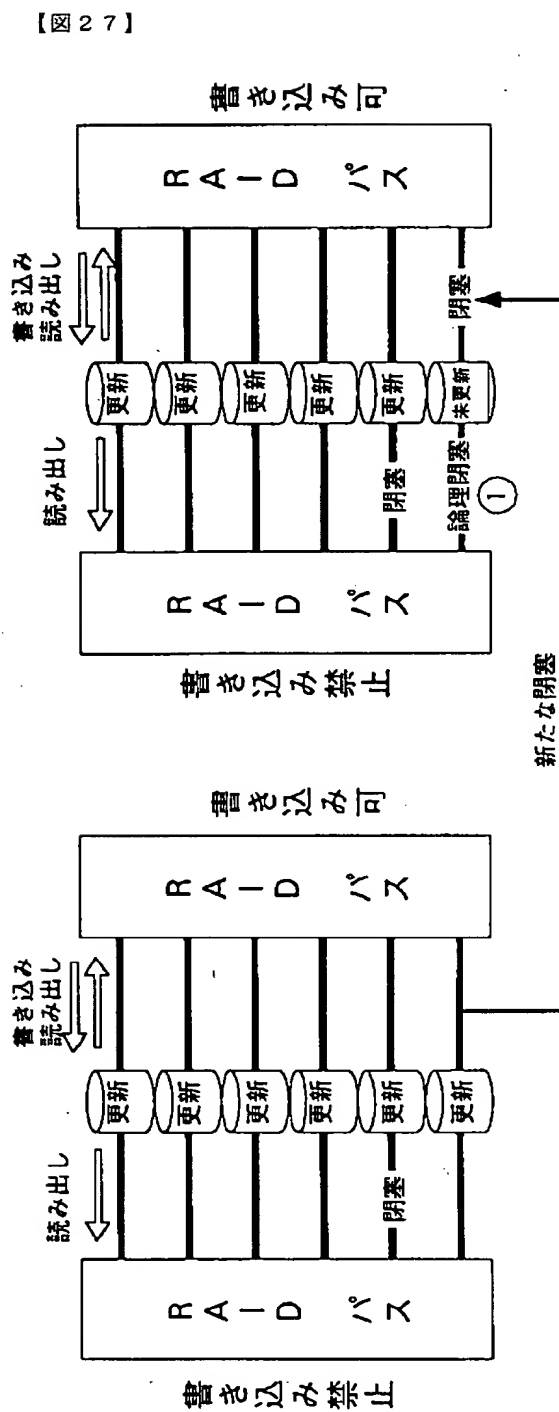


图27 入出力制限方式

読み出し専用/論理閉塞

ディスクパスに論理閉塞を設ける

ディスクバス障害 → 書き込み禁止のRAIDバスを構成するディスクバスを論理閉塞①



可 込 老 書

RAID パス

書き込み
読み出し

更新 更新 更新 更新 更新 未更新

しほみ跡

RAID パス

書き込み禁止

書き込み可

RA-D パス

書き込み
読み出し

更新 更新 更新 更新 更新 更新

幸み主一、

RAD パス

書き込み禁止

發

寒 閨 理 錦

器賦

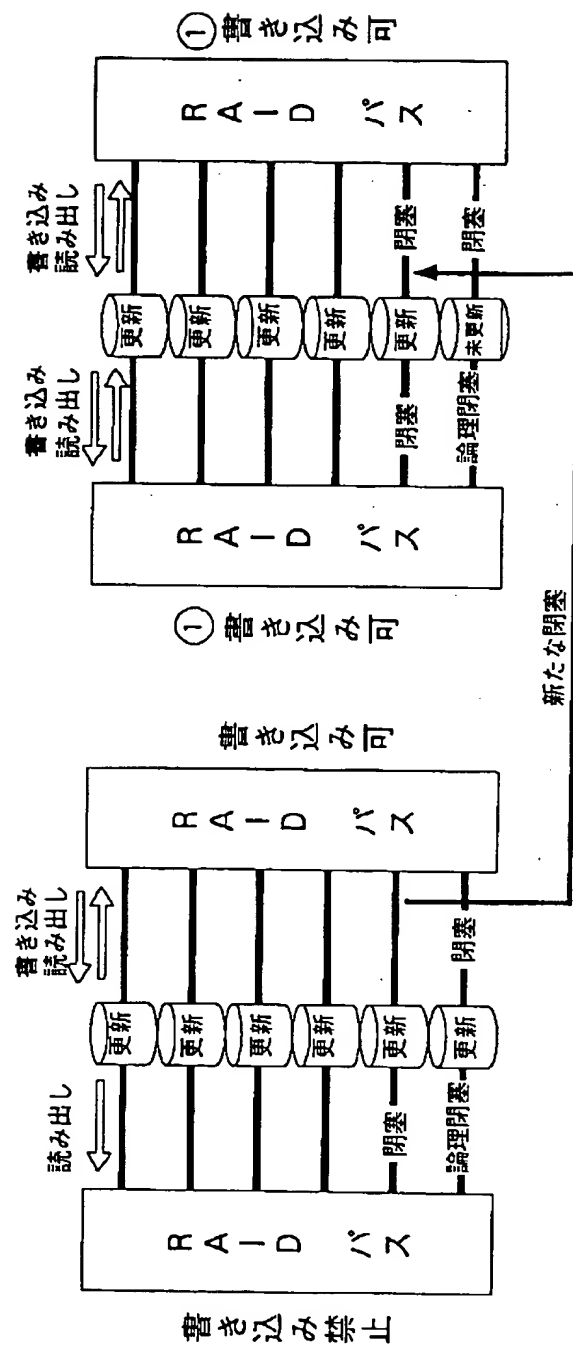
新たな閉塞

図28 入出力制限方式

両方のバスが閉塞の場合

入出力制限を書き込み可に戻す

ディスクバス障害 → 入出力不可能なディスク装置がRAIDバス間で一致するなら①
 両方書き込み可



【図28】

【図31】

図31 論理閉塞

クラスタ間で閉塞ディスクバスを一致させる

ディスクバス障害発生 → 他方のクラスタのディスクバスを論理閉塞 ①
 (例外、RAIDバスが閉塞なら論理閉塞しない)

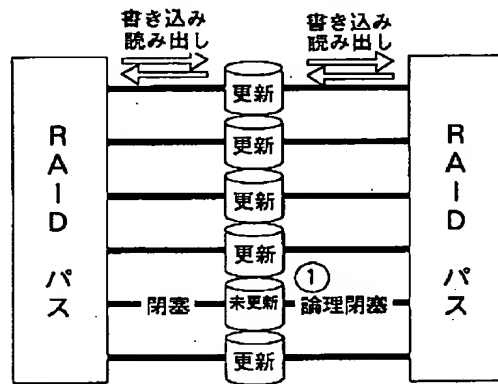
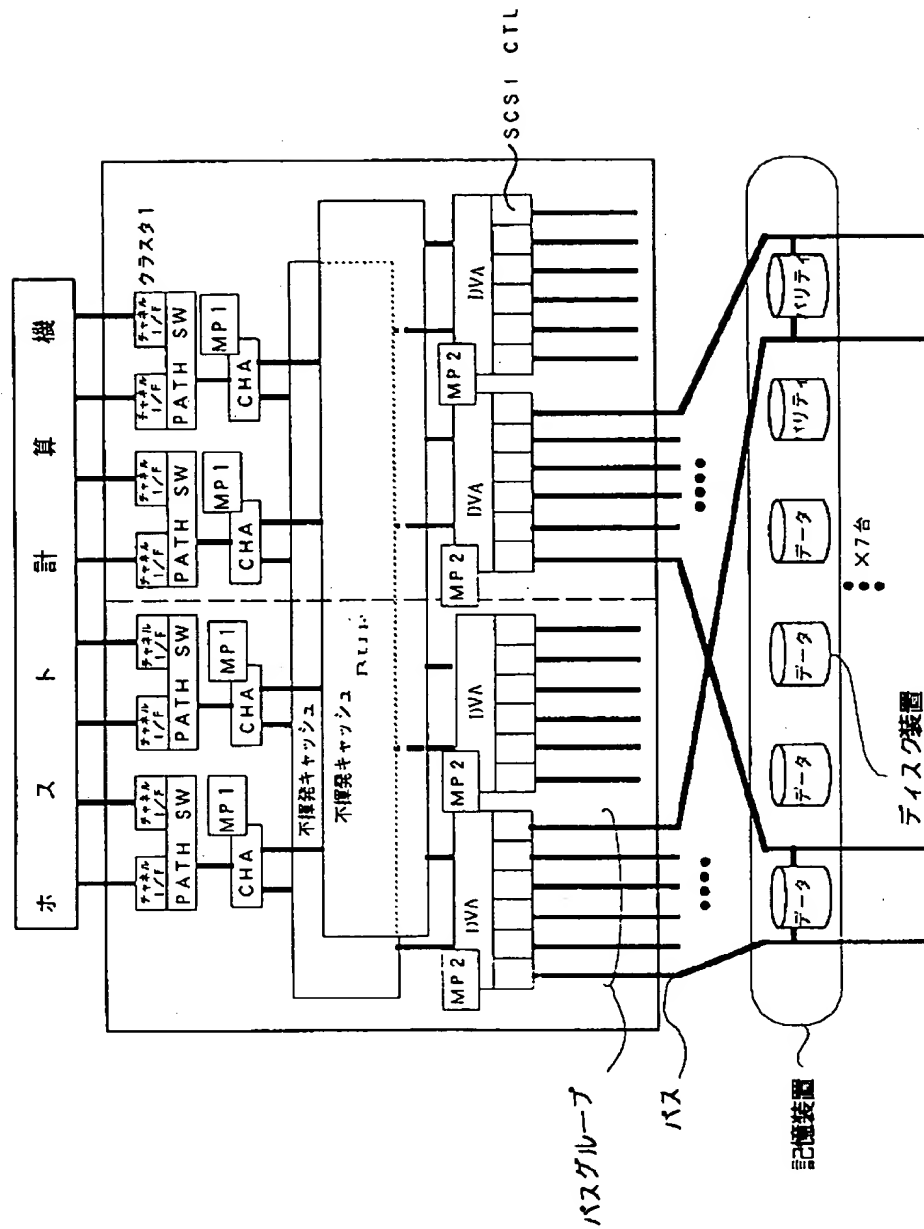


図32 ハードウェア構成図



(72)発明者 荒井 弘治
神奈川県川崎市麻生区王禅寺1099 株式会
社日立製作所システム開発研究所内